36462

National Library of Canada
Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* George MacLaggan PUGH

TITLE OF THESIS/*TITRE DE LA THÈSE* The Use of Psychological Tests for the Selection of Policemen

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Ph. D.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1978

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* O. O. Spearman

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* 25 April 78 SIGNED/*SIGNÉ* George Pugh

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* 120 Homestead Crescent
Edmonton Alberta
T5Y 2Y3

NL-91 (3-74)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

NL-339 (3/77)

THE UNIVERSITY OF ALBERTA

THE USE OF PSYCHOLOGICAL TESTS FOR THE SELECTION OF POLICEMEN

by

(C) GEORGE MACLAGGAN PUGH

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA
SPRING 1978

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "The Use Of Psychological Tests for the Selection of Policemen", submitted by George MacLaggan Pugh in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Supervisor

James C. Hackler

Paul F. Zelhart

External Examiner

Date April 25, 1978

TO MY PARENTS

MR. AND MRS. GERALD G. PUGH

Abstract

This research was directed towards developing a battery of psychological tests for selecting policemen. Accurate assessment of an applicant's ability to develop police job skills was considered to be essential in making decisions at the time of hiring. If a consistent relationship can be found between psychological test variables and job performance measures, knowledgeable predictions regarding the performance of policemen applicants may be made.

Two recruit classes of the City of Edmonton Police Department served as subjects for this research. The sixty-one (61) police recruits were tested within the first week of their training classes. The test battery included the following: Wonderlic Form B, Incomplete Sentence Blank, Draw-a-Person, Thematic Apperception Test, Strong Vocational Interest Blank, Cornell Index, Hostility and Direction of Hostility Questionnaire, Wechsler Bellevue Vocabulary, California Psychological Inventory, Situation Tests, and Three-Minute-Tests of expression.

For the projective tests, Situation Tests, and Three-Minute-Tests, scoring procedures were designed in accordance with the goals of the research. It was shown (with three reliability measures) that all of the scoring procedures designed for this research were applied in a reliable manner

by three raters, who were trained in these scoring techniques.

Three levels of police performance after two years on the job were ascertained by having the subjects rank-ordered in their squads by the sergeants of those squads. For reliability purposes a second performance appraisal was held one year following the first appraisal (first appraisal, March-April, 1976; second appraisal, March-April, 1977). To investige reliability even further, a third appraisal was completed in January-February 1978. During this appraisal period, as before, the sergeants' ratings were collected. Also, during this appraisal period, performance appraisals of the subjects were completed by senior constables and staff-sergeants. The performance measure was shown to be reliable.

Analysis of Variance and Multiple Discriminant Analysis statistical techniques showed that certain variables were predictive of police performance. Significant differences (Analysis of Variance) among the three groups occurred with the following variables in the directions predicted: Response Omitted, Incomplete Sentence Blank; Capacity for Status, California Psychological Inventory; Directions, Situation Test. The Multiple Discriminant Analysis statistical technique was used with most tests with four or more variables and also for the entire test battery. When used for the entire test battery only one subject was misclassified indicating very accurate predictability. Conclusions drawn from the two discriminant functions developed showed that high level policemen have interests in intellectual activities, high motivation for police work, and good ability to tolerate stress.

Acknowledgments

I wish to thank Dr. D. Spearman for all that he has contributed to this study. His efforts in guiding me towards completion of this project are very much appreciated.

I thank also the other members on my thesis committee; Dr. T. Nelson, Dr. P. Zelhart, Dr. B. Sinha, Dr. J. Hackler, and Dr. S. Smith. Their constructive comments were valued and helped in forming the final drafting of the thesis.

I would also like to thank the City of Edmonton Police Department for their help and cooperation. Special thanks to Chief R. Lunney, Deputy Chief G. Bishop, Superintendent J. Rodger, Superintendent D. McHarg, Inspector R. Foster, Inspector C. Ullrich, Staff-Sergeant U. Anders, and Constable C. McAdam.

Thanks to Mr. Berry Hsu and the Department of Psychology who made computer funds available for the large amount of statistical analyses that were required for this research.

I thank also my present employers, the Alberta Hospital Edmonton and Dr. K. Checkley, who cooperated so well in allowing many trips to the university during the final stages of the thesis. Their assistance and encouragement will not be forgotten.

Very special thanks to my wife Chris, whose encouragement from beginning to end never wavered, and whose presence in my life makes difficult tasks such as this so much easier.

TABLE OF CONTENTS

CHAPTER PAGE

I. INTRODUCTION 1

The Problem 1

Use of Psychological Tests 3

Problems in Conceptualizing the Good Policeman 8

Qualities of Good Policemen 8

Roles of Good Policemen 12

Concepts of Good Policemen 17

Developing Concepts of Good Policing 18

Summary: Concept of a Good Policeman 22

Local Attitudes Towards Policing 23

Different Approachs to Obtaining a Criterion 28

Problems with Research of this Type: Validity of Criterion 35

II LITERATURE REVIEW 41

Intellectual Level of Policemen 41

Interests and Aptitudes of Police Groups 51

Personality Characteristics of Policemen 64

Policemen and Stress 85

Situation Tests and Police Selection .. 86

Research Hypotheses 106

PAGE

II METHOD 111

Subjects 111

Description of Psychological Tests and Administration 114

Experimental Design 144

IV RESULTS 155

Wonderlic Form B 155

Incomplete Sentence Blank 156

Draw-a-Person 159

Thematic Apperception Test 164

Strong Vocational Interest Blank 164

Cornell Index 173

Hostility and Direction of Hostility Questionnaire 173

Wechsler Bellevue Vocabulary 175

California Psychological Inventory 175

Situation Tests and Three-Minute-Tests 178

General Analysis of Best Test Variables 180

V DISCUSSION 188

Intellectual Measures 188

Interest Measure 191

Personality Measures 194

PAGE

Multiple Discriminant Analysis of Test Battery 205

Concluding Remarks 210

***

REFERENCES 214

APPENDIX 1. POLICE JOB DESCRIPTIONS 245

APPENDIX 2. ATTITUDES OF EDMONTON RESIDENTS TOWARDS POLICEMEN 249

APPENDIX 3. POLICE STUDIES: CRITERIA USED 251

APPENDIX 4. INCOMPLETE SENTENCE BLANK 255

APPENDIX 5. SCORING FORM FOR THE INCOMPLETE SENTENCE BLANK 259

APPENDIX 6. SCORING FORM FOR THE DRAW-A-PERSON TEST 261

APPENDIX 7. SCORING FORM FOR THE THEMATIC APPERCEPTION TEST 267

APPENDIX 8. SCORING FORM FOR THE SITUATION TEST AND THE THREE-MINUTE-TEST 269

APPENDIX 9. INSTRUCTIONS AND RATING FORM FOR THE PERFORMANCE ASSESSMENT 272

APPENDIX 10. PROBLEMS WITH RELIABILITY OF PERFORMANCE RATINGS 275

APPENDIX 11. RATING BY SERGEANTS OF THREE WRITTEN INCIDENTS 284

APPENDIX 12. THE ATTRIBUTE RATING FORM GIVEN TO THE SERGEANTS 288

PAGE

LIST OF TABLES

Table 1: Use of Psychological Tests in the United States (1955-1972) 4

Table 2: Selection of Adjectives to Describe the Ideal Policeman by Police Officers and Students 26

Table 3: Early Intelligence Scores of Police Applicants and Policemen 42

Table 4: Mean Scores on the "Policeman Scale" as Reported by Johannson and Flint (1973) 61

Table 5: Interest Areas of Policemen 65

Table 6: Police Scores on the Edwards Personal Preference Schedule 71

Table 7: Test Performance of Success and Fail Candidates (from Mills, McDevitt, and Tonkin, 1966) 102

Table 8: Rank Order Correlations of Successful Candidates (from Mills, McDevitt, and Tonkin, 1966 104

Table 9: Selection of Recruits for Class 37 and Class 38 113

Table 10: Means and Correlations of Raters on "Overall Rating" scale for the Incomplete Sentence Blank 119

Table 11: Chi-Square Reliability Measures of the Incomplete Sentence Blank 122

PAGE

Table 12: Means and Correlations of Raters on "Overall Total" scale for the Draw-a-Person 125

Table 13: Correlations between Ratings on the six TAT slides 127

Table 14: Chi-Square Statistics for the TAT ... 128

Table 15: Correlation between ratings on the Situation Tests and Three-Minute-Tests 141

Table 16: Chi-Square Statistics for the Situation Tests and the Three-Minute-Tests 143

Table 17: Correlations Among Ranking Measures . 147

Table 18: Comparison of Correlation and Parital Correlation of Ranks 148

Table 19: Correlation Among Rating Measures ... 149

Table 20: Chi-Square Reliability Measures of the Three Incidents 151

Table 21: Most Important Attributes of Policework As Selected by Police Sergeants 153

Table 22: Group Means (Standard Deviations) and F Statistics for the Wonderlic .. 156

Table 23: Group Means (Standard Deviations) and F Statistics for the Incomplete Sentence Blank 157

Table 24: Prediction Results of Incomplete Sentence Blank 159

Table 25: Group Means (Standard Deviations) and F Statistics for the Draw-a-Person Test 160

PAGE

Table 26: Prediction Results of the Draw-a-Person Test 163

Table 27: Group Means (Standard Deviations) and F Statistics for the TAT 165

Table 28: Group Means (Standard Deviations) and F Statistics for the Strong Vocational Interest Blank 166

Table 29: Prediction Results of the Strong Vocational Interest Blank 171

Table 30: Order of Variables Selected in SVIB Multiple Discriminant Analysis 172

Table 31: Group Means (Standard Deviations) and F Statistic for the Cornell Index 174

Table 32: Group Means (Standard Deviations) and F Statistics for the Hostility and Direction of Hostility Questionnaire 174

Table 33: Group Means (Standard Deviations) and F Statistic for the Wechsler-Bellevue Vocabulary 176

Table 34: Group Means (Standard Deviations) and F Statistics for the California Psychological Inventory 177

Table 35: Prediction Results of the California Psychological Inventory .. 178

Table 36: Group Means (Standard Deviations) and F Statistics for the Situation Tests and the Three-Minute-Tests 179

Table 37: Prediction Results of the Situation Tests and the Three-Minute-Tests 181

PAGE

Table 38: Prediction Results for the Best Variables of the Test Battery 183

Table 39: Order of Variables Selected in Test Battery Discriminant Analysis 184

Table 40: Standardized Discriminant Function Coefficients for the two Discriminant Functions 187

Table 41: High Positive and High Negative Weights for Function 1 206

Table 42: High Positive and High Negative Negative Weights for Function 2 209

Table 43: Ratings by Sergeants of the Three Written Incidents 287

List of Figures

PAGE

Figure 1: Plotting of Group Centroids 185

CHAPTER I

INTRODUCTION

The Problem

The purpose of this research is to develop and validate a battery of psychological tests that will assist in the selection of policemen. The selection of policemen is often based on characteristics assumed to be related to good job performance. The actual relationships between characteristics of the applicant and later performance on the job, however, may not have been well substantiated. This research, if successful in its intent, will examine and, if possible, validate, relationships between characteristics of police applicants and job performance.

The procedure for selecting policemen may be improved with the use of the following approach. The relationship between certain characteristics of the applicants and later measures of job performance may be examined. For example, variables used for predicting the job performance of a future policeman, such as the job interview, scores on entrance tests, school records, background investigation findings, and psychological test results, can be placed for statistical purposes on a numerical scale. Similarly, job performance criteria, such as annual performance ratings, commendations, and merit awards, with which these predictors should be related can also be scored in a numerical fashion. Once this categorizing approach has been implemented the relationships between the predictors and the criteria can be examined closely and suggestions for improving these relationships may be considered. Improvement in the selection process may take many forms. Present test variables may be found to be unrelated to future job performance and therefore discarded. New selection tests may be added and their relationship with job performance measures ascertained through statistical analysis. The criteria can be changed so that a more accurate representation of job performance results. Through these approaches to understanding the relationship between selection screening devices and job performance measures a more refined selection process may be developed.

Within this study the selection of police recruits for the City of Edmonton Police Department will be considered. A previous study (Ferguson, 1974), also conducted in cooperation with the City of Edmonton Police Department, was primarily concerned with quantifying and examining the relationship between the predictor variables used at that time and criteria variables. The present research focuses on the problem of developing a battery of psychological tests to be used in police selection procedures. This problem will be concerned directly with attempting to predict future job performance of police applicants from psychological test variables. For this reason the studies to be reviewed will be related to the use of psychological tests and police selection.

Use of Psychological Tests

Few studies concerning the use of psychological tests in police departments in Canada have been reported in the literature. Because of the differences between and within countries, the conclusions and generalizations drawn from all police studies have to be made with caution when considering their applicability to each police department.

It is reported that psychological tests and psychologists are being included to a greater extent in the selection procedures of police departments in the United States. As shown in Table 1 there are four studies which surveyed the types of testing used by police departments in the United States. Frost (1955) sent questionnaires to 33 centers with a population over 150,000. From the 25 cities which responded the author reported that no tests of a psychological nature were being used. Oglesby (1957) found that only 29 percent of the cities he surveyed (with populations over 100,000) used psychological tests. Narrol

Table 1

Use of Psychological Tests
in the United States (1955-1972)

STUDY	NUMBER OF CITIES SURVEYED	CITIES OVER	REPLIES	PSYCHOLOGICAL TESTS USED	STANDARDIZED TESTS USED
Frost (1955)	33	150,000	25	0%	0%
Oglesby (1957)	111	100,000	90	29%	?
Narrol & Levitt (1963)	61	150,000	55	100%	42%
Murphy (1972)	307	50,000	203	39%	100%

and Levitt (1963) six years later found that all of the cities surveyed (with populations over 150,000) used some form of psychological test. They noted that 42 percent used a standardized intelligence test while 87 percent used a nonstandardized police selection test. Only sixteen percent used a standardized personality test. (Standardization implies that a uniformity of procedure in administering and scoring the tests has been provided by the test constructor along with norms for the interpretation of the test). Murphy (1972) surveyed police agencies serving 50,000 or more people and having at least one hundred police officers in their force. From those replying he found that 39 percent used psychological tests to evaluate potential law enforcement officers.

Information from these survey studies is inconsistent. The reported large differences between 1955 (0%) and 1963 (100%) indicate that conclusions drawn may be misleading. These studies, however, have indicated that some police departments are not using standardized tests. (Therefore test score distributions for representative samples drawn from the population are not known and references points for comparison with the individuals being tested would not be available.) If tests used for police selection are not standardized or have not been validated in terms of predicting the future job performance of policemen the meaningfulness of the test results would be unclear.

Many psychological tests that have been included in police selection test batteries were originally designed for the assessment of mental health problems within a hospital setting. This clinical environment is different from a selection-assessment setting. The former tends to encourage honesty and openness, while the latter encourages competitiveness and good self-presentation. Murphy (1972) has pointed out that the Minnesota Multiphasic Personality Inventory (MMPI), a test designed for a diagnostic or theraputic setting, was the test most frequently used by police departments. Reiner, Setzer and McCormack (1960) have discouraged the use of this test for the selecting of police recruits. They have indicated that when it is used for personnel selection the possibility of faking the test in a favorable direction is much greater then when it is used in the clinical situation for which it was originally designed. Anastasi (1968) discusses this problem and has the opinion that: "As long as a subject has sufficient education to enable him to answer a personality inventory ... he probably has the ability to alter his score appreciably in the desired direction" (p. 456). These objections are directed towards the validity of the psychological tests used, i.e., how effective is the test in predicting an individual's behavior in specified situations, for example future job performance as policemen. Very little substantial data showing the validity of psychological tests when used for police selection are available. It is possible that candidates are being selected on the basis of test results (which are usually used in a clinical setting) without a clear understanding of the relationship between these test results and occupational success. If psychological tests are to be used it is important that police selection officers are aware of the research that demonstrates relationships between test variables and job performance.

It is upon these two basic elements, job performance measures and psychological test variables, and the relationship between them, that the following review of the literature will focus. First, the important aspects of job performance will be considered. These important aspects will be clarified by examining the concept of a "good" policeman. This is essential so that new recruits will be assessed on the important dimensions required for policework. A discussion of the criterion problem, that is, what form of criterion should be used, will also be examined. A consideration of the criterion selected is necessary because its selection places limitations on the generalizations and conclusions drawn from the research.

Second, the more formal literature review will be presented. This will include psychological test scores of police populations in areas related to intelligence, interests and aptitudes, and personality. The possibilities of assessing recruits in job related situations will also be examined. The relationships of these psychological test variables and a multitude of job performance criteria that have been found in past research will be considered carefully in designing the present research project. From this approach hypotheses concerning the relating of psychological test variables to job performance measures may be stated.

Problems in Conceptualizing the Good Policeman

An important prequisite in this research was the careful consideration given to the nature and actions of the good policeman. Terms used to describe "good" policing were reviewed and organized for classification. This base was to serve both in the selection of assessment techniques given to applicants for police work, and in the evaluation of their performance at later times on the job.

It was necessary, then, to develop an understanding of policing. In reviewing the literature towards this end, it was noted that qualities of policemen were presented, different roles of policemen were discussed and, finally, some concepts of policing were suggested. These three categories were judged to be useful ways of organizing the presentations of the many authors writing on the subject, and are expanded below.

Qualities of Good Policemen

Job descriptions. "Qualities" may be defined as an important aspect of a thing that makes it what it is (Webster Encylopedic Dictionary, 1973). Certain qualities (terms descriptive of behavior) are involved in good police function. The qualities required in policing vary according to the circumstances. Job descriptions were helpful in listing the essential qualities that recruiting officers look for in applicants for police work. Holmes (1942) and Baehr, Furcon, and Froemel (1968) developed job descriptions for policemen. (See Appendix 1, page 245, for a listing of these job descriptions.) From the qualities they listed (Holmes listed eleven, Baehr et al. listed twenty) the following have been selected as important and also susceptible to measurement within this research.

(1) React quickly and effectively to problem situations.

(2) Exhibit initiaive, problem solving capacity, effective judgment, and imagination in coping with complex situations.

(3) Demonstrate mature judgment (commonsense) in deciding to make an arrest, give a warning, or use force.

(4) Tolerate stress in a multitude of forms.

(5) Maintain a balanced perspective in the face of constant exposure to the worst side of human nature.

These qualities will be examined in more detail in a later section of the literature review. They will be considered in terms of how well they can be assessed with the use of certain psychological tests (e.g., situation tests, problem-solving tests).

Police views of important police qualities. Another approach that has been used in deciding upon the nature of the policeman's job has been to ask policemen to select the qualities essential for those in their position. Stirling

(1972) had policemen with 18 months experience select from forty qualities ten which they believed essential in good policemen. They selected: (1) "commonsense", (2) "alertness", (3) "job knowledge", (4) "well trained", (5) "honesty", (6) "intelligence", (7) "responsibility", (8) "dedication", (9) "reliability", (10) "appearance". Stirling commented on the first attribute. "Whenever something goes wrong for a policeman, the most frequently heard explanation for the occurrence is that 'he didn't have common sense'" (p. 239). After six months further job experience these same policemen ranked "courtesy" within the top five attributes.

Other authors have used a similar approach in having policemen rank important qualites required for their work. Westley (1951) noted "alertness" and "competence" as being highly rated. Finney (1972) found the attributes of "honesty", "responsibility", and "self reliance" to be considered among the most important by the police group he surveyed. A police instructors group (Stirling, 1972) gave "honesty", "commonsense", and "understanding" as important qualities. Devine (1972) stated that men with six months police experience found "commonsense" and "honesty" as the most essential qualities required for police work.

From these studies it appears that "commonsense" is the quality policemen most often considered essential for their position. This quality is included among the essentials in the job description by Baehr and her colleagues (1968). For example, that job description, as mentioned earlier, included the following skills: reacting "quickly and effectively to problem situations", exhibiting "problem solving ability", demonstrating "mature judgment", "acting effectively in extremely divergent interpersonal situations". All of these functions could be included within an operational definition of "commonsense".

When considering important qualities (similarly with roles and concepts) in assessing candidates for the position of policemen the difficutly with which those qualities can be measured has to be taken into account. For example, "commonsense" appears to be important but how accessible is it to measurement? Is it possible to separate those who have it from those who do not? This dimension of measurability has to be considered when choosing those qualities, roles, and concepts to be examined in the assessment.

For the purposes of this study "commonsense" as it relates to intellectual development was considered measurable. Similarly, the measuring of ability to handle stress in interpersonal situations may be assessed with the use of "situational tests". A review of tests of intellectual development, and situation tests as they relate to police recruitment will follow later in the research.

Qualities required for policing are different than roles. Roles attempt to assemble certain qualities and help in directing their application. The previously presented essential qualities required for policemen were not accompanied by role directives orienting their suitable application in policing.

Roles of Policemen

Roles have been defined by Yinger (1965) as a "unit of culture referring to the rights and duties, or normatively approved patterns of behavior" for the occupants of a given position. A role is the part a person plays in a situation. In police work, it is a part played in real life. There are a number of roles policemen must fulfill. Indeed, conflict among roles may exist. Yinger (1965, p.74) defined four types of role conflict.

> 1. Internal role conflict occurs when an individual has internalized a role that includes contradictory expectations or when he occupies two or more positions that carry incompatible role expectations.
>
> 2. External role conflict occurs when an individual is confronted with incompatible expectations from two or more persons in his position network or networks.
>
> 3. Extrarole conflict occurs when an individual perceives that others hold different expectations for him as an incumbent of a single position.
>
> 4. Interrole conflict occurs when an individual perceives that others hold different expectations of him as the incumbent of two or more positions.

The various categories of role conflict result in problem areas for individual policemen. Yinger (1965) has also described the "internalized role", which refers to that part of the self which represents an individual's tendency to perform a role in a certain way. For this research, the "internalized role" will be discussed more fully under the topic concept of policing. For the moment, it may be said that the concept of policing possessed by the individual results in certain priorities being placed upon role selection and tendencies to resolve role conflicts in certain directions. For example, if a role conflict exists between cautioning a traffic violator and giving him a ticket (policeman as public instructor, versus policeman as law enforcer) the policeman's choice is dependent upon his concept of policing. It will be noted that roles and concepts are almost inseparable; nevertheless, it is likely the former are directed by and serve the latter. Before outling concepts of policing a brief consideration of general roles of policing will be examined.

The role of law enforcer. Policemen have as one of their main functions the enforcement of legal norms. Westley (1970) writes that "law enforcement is the legal job of the police" (p. 140). Many writers in the police area note that policemen themselves view law enforcement as their major responsibility. The police function required results in "identifying those thought to be guilty of having committed a criminal offense and subsequently proceeding against them" (Weiner, 1976, p. 10).

The role of mainiaining social order. (See Alex, 1969; Reiss, 1971; Silver, 1967; Skolnick, 1966; Westley, 1970; Wilson, 1968.) This role involves police behavioral qualities such as dealing with noisy neighbors, loiters, and drunks. Weiner (1976) has pointed out the following.

> That while enforcing the law is a difficult task it is a fairly unambiguous one because it involves clear-cut acts such as murder, speeding, and theft. On the other hand, maintaining order is a much more difficult task because it involves an undefined condition: public order. Therefore the law is one resource among many that the patrolman may use to deal with disorder. (Weiner, 1976, p. 11)

Order maintenance is unpredictable, (and requires greater discretion (e.g., to arrest, or not to arrest). Wilson (1968) and Cummings, et al. (1965) reported that most police work was associated with order maintenance.

Policemen must be careful that order does not become an overriding concern to the point where the sensible law is ignored. This is typical in totalitarian states where civil rights are ignored so that order can be maintained. The concept behind the role of maintaining social order may be judged that of social harmony. This concept may be difficult for some policemen to define and hold in real, immediate, and future perspectives.

The role of public-servant. Policemen are also required to offer community services such as helping the unfortunate, giving directions, ambulance and funeral escorts, first aid, advice-giving, recovering lost persons and property. Serving the public is an important role for policemen because within this role they relate to citizens in a nonpunitive, helping manner. The concept behind this role is probably that of providing humanitarian and community service. The lower socio-economic classes more often require police services in these areas then do other citizen groupings. Cummings, et al. (1965) and Bittner (1970) have suggested four reasons why policemen are used in this role capacity. First, they provide a 24 hour reliable service that other organizations do not provide. Second, social services to this segment of the population are lacking and policemen are visibly available. Third, when social services are available, people are not aware of them. Fourth, policemen have a unique ability when intervening, "they can employ a 'non-negotiable coercive force' which need not be explained to anyone and which, practically speaking, cannot be opposed by anyone" (Weiner, 1976, p. 12).

In fulfilling this public-servant role, law enforcement itself is seldom used but its potential use is always present. The policeman's special access to law enforcement gives him a large advantage over other professionals (counsellors, social workers) who may also attempt the public-servant role.

> As the symbolic backdrop of the police licence and capability is always present whenever a policeman responds to a problem, he is always responding as a policeman and not as a social worker, whether amateur or professional. Indeed the continual presence of the police licence and capability mitigates against him ever being able to play the role of a social worker as everyone (including the policeman) will know that ultimately he has access to the means uniquely accessible to policemen. (Shearing and Leon, 1977, p. 342)

The above roles have been chosen as important within this study. For further information regarding other roles of policemen (role of value-model, role of buffer between people and power structures, role of policemen in measuring social tolerance) see Wolfgang (1968), Black (1968), and Weiner (1976).

It may be that there is an incompatibility among some of these police roles. Also, some of these role behaviors may not be measurable. Other role behaviros may be more accessible to measurement in a long term study.

Within this research two roles of policing will be focused upon for assessment purposes. A person's ability to acquire the roles of the public servant and law enforcer may be reflected in measures of self-concept and concept of others. These measures may be derived from certain projective psychological tests. These will be considered in detail later in the literature review.

Concepts of Policing

For the present application, a concept may be defined as the idea, principle, value, or ethic serving to direct an action or role. Concepts held by the policeman influence role formation and selection. The subtle ordering, emphasizing, and priority-giving of these police roles varies according to the situation and the concept of policing that the individual holds.

Contributing to the development of different policing concepts is the ease with which some police roles may be adopted in preference to other more difficult police roles. In these instances the role dominates the concept when, rather, it should be the reverse. Through rationalization, new concepts of lesser worth may emerge. For example, police work in general is simplified when the policeman's expected behavior is unambiguous (e.g., the law enforcement role). The action to take in pursuing a murder conviction is more obvious than when intervening in a marital dispute. For clear and simple situations little possibility of role conflict exists. For murder, policemen see themselves as law enforcers. For other situations, such as the marital dispute, role conflicts are likely (e.g., policeman as enforcer versus policeman as counselor or public servant). It is easy to understand why policemen prefer the role of law enforcer because of the relative clarity in defining appropriate action. The concepts held determine the perceptions of a situation. Where skills and roles are appropriate they are enacted. Where they are not, other concepts may be activated. Good policemen have the right concepts and can select appropriate role behaviors.

The concept of a "good" policeman for this research was outlined as follows. The good policeman has the values, principles and ability to appropriately select, blend, and fulfill the different roles of policing (law enforcement, maintaining order, public servant) according to the demands of the job situation. He has the qualities and skills to enact each role. In general the roles have been listed in order of importance but, as mentioned previously, they are highly dependent upon situational determinants and individual concepts of policing.

It may be suggested that the "good" policeman should understand which role is most appropriate in each situation. For this research, assessment of police candidates will be directed towards acquiring policemen that have, or have the potential to acquire; the principles, actions, and qualities to fulfill the major roles of policing.

Developing Concepts of Good Policing

The selection of the appropriate police role for varying situations is dependent upon the concept of policing possessed by the individual. From a review of police philosophy, especially Peel's principles of law enforcement, two main concepts of policing emerge as important in the development of "good" policemen.

(1) Conception of police work as an effort to improve the welfare of the community (social harmony). The initial step in this concept is put into action through police work to prevent crime and disorder (Peel's first principle). Actions in this regard may involve a number of different police role behaviors: here the public servant role in the preventive aspects is often used. Bittner (1970) has elaborated this concept in describing the development of policing.

> The sentiment that could not abide the more archaic forms of repressive control of deviance and disorder was an expression of cultural and idealogical change initiated in the nineteenth century. It is best described as the rise of the sustained, and this far not abandoned, aspirations of Western society to abolish violence and install peace as a stable and permanent condition of everyday life. (p. 16)

Also serving this concept is police action that involves taking part in educational processes so that appropriate law-abiding attitudes are developed by citizens. With these efforts the policeman helps develop the willing cooperation of the public in voluntary observance of the law, Peel's third principle.

The focal point this concept of a "good" policeman is police action to benefit the well-being of the community. Preventing crime, deterring crime, and finally law-enforcement are the action equivalents of the concept of benefiting the community. If this concept is adopted by a police department and carried out by its policemen, the example may persuade community endeavors of a similar nature. This concept may be measurable with the use of projective personality tests that assess attitudes towards self, community, and harmony within the community.

(2) Conception of high respect for individual rights. Another basic concept required for "good" police action is respect for the rights, worth, and dignity of individuals. It is a central democratic principle. Policemen demonstrate this concept by impartially enforcing the law, and giving service.

This principle, with which most policemen would agree, may be very difficult to implement. The policeman's maturity in handling unruly and obnoxious individuals and his tolerance of verbal abuse will demonstrate his ability to put this principle into action. For example, a professional policeman realizes the difference between law enforcement and the role of the courts and therefore continually demonstrates his respect for an individual's rights. Consider what happens when policemen go beyond their legal duties.

> I have known officers who were mad at every suspect they ever arrested. These officers were actually angry because of what the suspect had done, said, or looked like. This unprofessional attitude breeds resistance and hostility making the police job more difficult, if not impossible.
>
> ...The quality we look for here (in professional policemen) is equanimity. The professional policeman does not demonstrate emotional involvement in his work to the public. He does not demonstrate anger to the offender; he does not go into a rage when verbally abused. He keeps his cool. (Eldefonson, 1973, p. 119-120)

The professional quality of "equanimity" may be directly measurable on personality tests. These types of test measures will be discussed later in the literature review.

An important quality required to interact with citizens and demonstrate respect for individual rights is that of verbal skills. For example, the role of the public servant as mediator in disputes demonstrates the necessity of the high development of verbal qualities required in policemen.

> Policing demanded eloquence. If the twenty-eight young policemen agreed on any one point about their work it was the vital importance of verbal facility in every aspect of their jobs. "The crux of police work is that ability to talk to people", they would say. There were many uses of eloquence---it was the key to taking charge in public... Lacking a ready and capable ability to talk, an individual policeman simply could not make professional responses and accept the risks inherent in them. (Muir, 1977, p. 22)

The measuring of verbal skills under stress will be considered for this research.

In summary two important concepts are required for "good" police work. They are central to appropriate police role selection. The first is: to maintain and improve the welfare of the community through law enforcement and preventing crime. Second; the establishment and maintenance of good democratic relationships with the public. Respect from the public is as much engendered by the professional policeman's respecting the rights of citizens, as by his skilled actions of enforcement. The assessment of an individual's attitudes, personality traits, and verbal skills, with regard to these concepts will be considered within this research.

Summary: Concept of a "Good" Policeman

The "good" policeman is capable of filling the major role requirements of police work. The roles are guided by concepts of policing: namely, a concern with benefiting the community and a respect for the rights of individuals. The policeman must also be at a level of intelligence and maturity to acquire the more constructive values and concepts and thus to be able to choose the appropriate roles for the appropriate situation. Specifically, these requirements necessitate a mature development so that a 'mature perspective towards citizens' may be adopted. Finally, verbal skills are necessary. These are the basis requirements of a "good" policeman.

Local Attitudes towards Policing

Police department officials often strive to work cooperatively with law-abiding citizens. Knowledge of the attitudes of local citizens towards policemen may be helpful in discerning a general view of the job performance of the police department. Knowledge of citizens' concepts of policing may also be helpful in designing job performance measures.

Two studies have recently been completed regarding citizens attitudes towards local policing. In an unpublished study (Pugh, 1975) policemen and citizens were asked to select from three hundred adjectives of the Adjective Checklist (Gough, 1952) qualities that would be "descriptive of the ideal policemen". Eleven groups were tested. Seven of the groups were from the police department and were comprised of 17 police officers, 33 staff sergeants, 58 detectives, 51 sergeants, 13 police constables with 10 years experience, 22 police constables with five years experience, and 37 police recruits. Three of the groups were from the community and consisted of nine senior citizens, 14 summer school students (from a psychology class), and 12 members of a driver retraining group (required because of traffic violations to take the course). The remaining group was made up of 16 jail prisoners.

Among the groups within the police department a basic concensus appeared. It was observed that the adjectives selected centered around three points: relations with others ("reasonable", "understanding", "loyal", "responsible"); inner personality qualities ("mature", "stable", "confident"); and mental attributes ("clear thinking", "flexibility", "intelligent", "adaptable"). These results were consistent with other studies but the large number of adjectives from which to choose allowed the subjects greater specificity in selecting the adjectives. For example "commonsense" a quality not on this checklist could possibly include the qualities "reasonable", "understanding", "responsibility", and "stability".

In comparing citizens groups and the prisoner group with the policemen very few differences in concepts of the ideal policemen were noted. For example the adjectives selected by 70% to 80% of the prisoners were: "active", "calm", "dependable", "mature", "reliable", "alert", "capable", "civilized", and "friendly". All of these adjectives were selected by police officers at the over 90% level except for "friendly" which was selected at the 80% level of agreement. It appears that the prisoners' views of the ideal policeman are included within the police officers' viewpoints.

When the university students were compared with the police officers again no significant differences occurred.

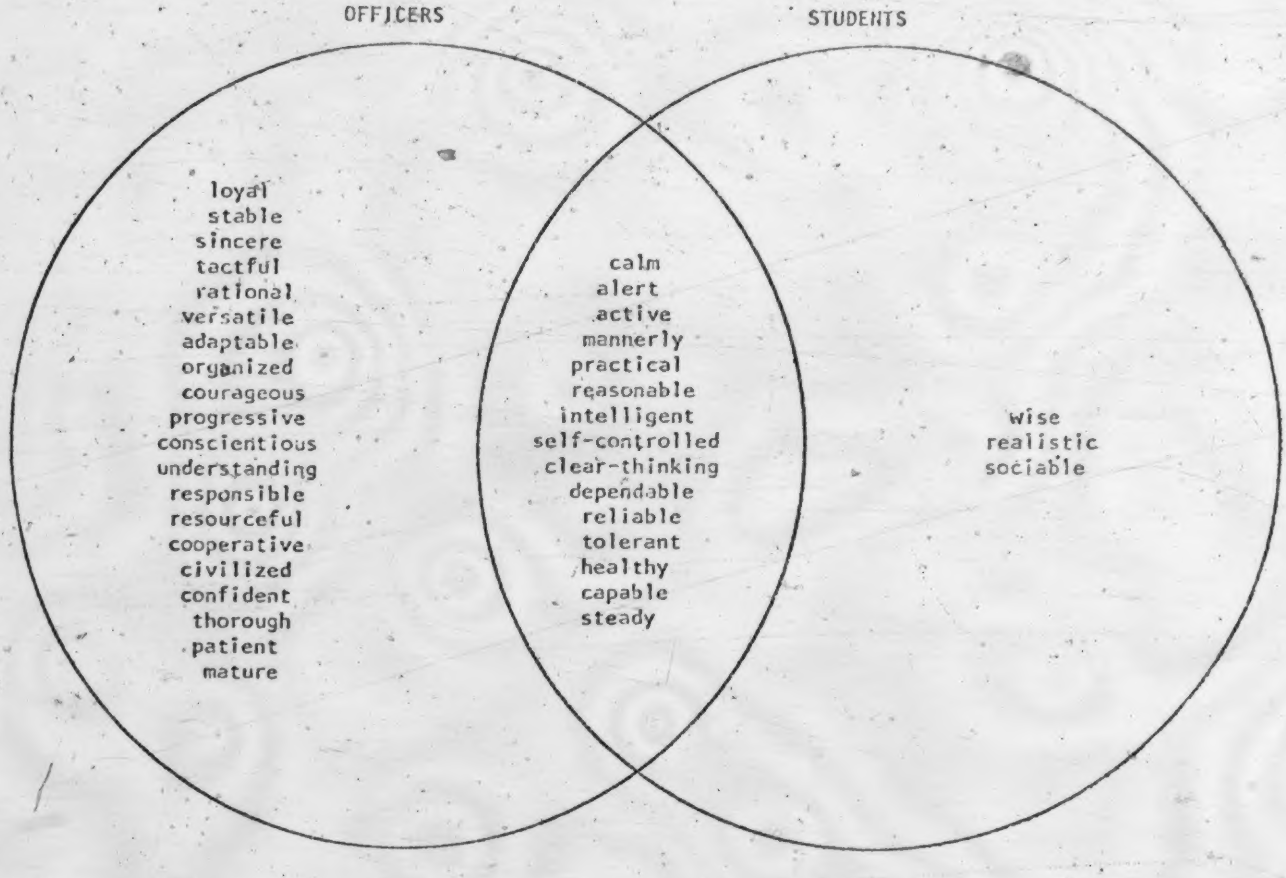
As shown in Table 2 there is a great amount of overlap between the students and the officers (the officers were considered representative of the policeman in general). Every adjective selected by over 90% of the students was selected by over 90% of the police officers except for three adjectives ("realistic", "sociable", and "wise") which were selected at the 80% level of agreement by the officers. Even the large group of adjectives that were selected by the police officers at the over 90% level of agreement were selected by the students to the degree that the frequency of the selections could not be considered to be significantly different.

Similar agreement between police officers, senior citizens, and the driver retraining groups were found. It was concluded in this study that the concepts of the ideal policeman held by policemen, citizens, and prisoners were not contrasting but similar in nature. However, policemen did consistently select more adjectives in their descriptions of the ideal policemen.

There is another local study that also examines citizens' views of policemen. The Alberta Bureau of Statistics and the Edmonton Police Commission (August, 1976) reported attitudes of citizens of Edmonton towards city policemen. This was a mail-out survey which was sent to 991 individuals randomly drawn from listings of telephone subscribers. Following the initial send-out two follow up

Table 2

Comparison of selections of officers and university students in 90-100% agreement range


OFFICERS
STUDENTS
loyal
stable
sincere
tactful
rational
versatile
adaptable
organized
courageous
progressive
conscientious
understanding
responsible
resourceful
cooperative
civilized
confident
thorough
patient
mature
calm
alert
active
mannerly
practical
reasonable
intelligent
self-controlled
clear-thinking
dependable
reliable
tolerant
healthy
capable
steady
wise
realistic
sociable

questionnaires were mailed. The final response rate was 86.7% (859 citizens).

Generally respondents to the survey "expressed highly positive attitudes toward the city police" (p. 1). For example, when citizens were asked: "In general, how favorable do you feel toward the police in this city?", 49.0% indicated "Very Favorable", 37.0% chose "Moderately Favorable", 8.0% chose "Slightly Favorable", 2.9% chose "Slightly Unfavorable", 2.0% chose "Moderately Unfavorable", and 1.1% chose "Very Unfavorable". When these results are condensed it is demonstrated that 94% of the citizens chose the favorable side of the response continuum while 6% chose the unfavorable side. It appears from the results that citizens of Edmonton have favorable feelings towards their policemen.

Responses to policemen on other questions also appeared favorable. Citizens rated the overall job performance of the city police as "Very Good" 51.5%, "Moderately Good" 49.7%, "Slightly Good" 6.8%, "Slightly Bad" 1.3%, "Moderately Bad" 0.6%, and "Very Bad" 0.1%. When these results are condensed it is shown that 98.0% selected the "Good" side of the response continuum while 2.0% selected the "Bad" side. A summary of this questionnaire is shown in Appendix 2, page 249.

The above two studies reflect citizen attitudes that suggest agreement with present policing policies. No significant conflict between citizens' attitudes and conceptions of policemen and those of the police department policy appeared evident.

Different Approachs to Obtaining a Criterion

The selection of job performance measures is extremely important in studies of this nature. There are many different approachs to solving criterion problems. Appendix 3, page 251, categorizes examples of criteria that have been used in other police studies. These studies have been grouped according to certain characteristics. These groupings will be pointed out with the problems relevant to each criterion type. This approach will aid in the selecting of good job performance measures for this research.

Performance ratings by supervisors as a criterion.

According to the literature review this type of criterion is the most frequently used. Problems with performance rating systems as they apply to this research have been outlined in the literature (Ghiselli and Brown, 1948, Thorndike and Hagan, 1969). These problems will be described briefly below.

> (1) The most important consideration is the extent to which the supervisor is aware of the performance of the person being rated. The supervisor himself has his own work to do and this may limit the amount of time he has to observe the ratee. Supervisors must have the time to observe and become aware of those they are rating.

> (2) The rater's concept of "good" job performance may be different from those who require the ratings. This problem may be alleviated by clearly defining the qualities that are evident in "good" and "poor" job performance.
>
> (3) The social relations between supervisors and workers will influence ratings. A supervisor shall tend to rate friends higher. If rating supervisors are made aware of the biasing potential of friendship they then may attempt to give a more accurate job performance score for all those they are rating.
>
> (4) The covertness and abstractness of the trait being rated and the rater's understanding of that trait will affect the accuracy of the ratings. If the trait being rated is an internal quality (e.g., feelings of insecurity) it will be more difficult to assess. Similarly if the trait is very abstract (e.g., initiative) each raters' definition of the term may vary. Raters should be instructed concerning the meaning of the terms upon which others will be judged before they complete any ratings. For example, a rater should have a clear understanding of the department's definition of terms such as "average performance", "above average performance", and "outstanding performance".

The above potentially invalidating aspects of a rating system have to be watched for and appropriate precautions have to be taken to minimize any detrimental effects on the accuracy of the ratings.

The supervisors' rating approach is presently being used in the police department where this research project was conducted. It requires sergeants to rate constables on 16 traits and give a written account for "Leadership", "Factual Details", and a "Narrative Assessment". The sergeant then discusses these results with the constable involved so that deficiencies can be pointed to as areas for improvement.

Ferguson (1974) in a study involving this same police department, factor analysed the supervisors' rating form. It was discovered that one factor was the best representation of the entire form. This factor was comprized mainly of the "Narrative Assessment" variable (loading of +.80). This could be described as a "Summary of Overall Performance" measure.

Another form of a supervisor's rating approach is the paired comparison method of measuring the job performance of an employee, wherein, a supervisor rating his men would compare each man with every other man. In this type of rating the supervisor would rate his men two at a time. For each possible pairing of men he would indicate which of the pair was the better. The main disadvantage of this system is that the number of ratings the supervisors have to complete is highly increased. For example if 10 men had to be rated, with this system 45 pairs of men would have to be considered. Another serious problem is that pointed out by Tversky (1969). It is possible that when assessing complex behaviors, paired comparison ratings may be inconsistent (intransitive).

An additional form of supervisors' ratings involves the measurement of preselected job performance variables (marksmenship, number of arrests, miles per contact with radar, number of commendations, absences, complaints, sick leave, etc.) as the criteria. This method, which usually involves objectifying the criteria, has been used in many research studies but not to the extent of the performance ratings. This approach may also be used if a quality is considered essential in policemen. For example, if the quality of "courtesy" could be objectified and made observable, it could stand as a criterion measure representative of job performance. If "courtesy" could be adequately assessed then policemen outstanding in this area would more likely be promoted.

When using this approach each preselected variable which serves as a measure of job performance has to be considered on its own merit. A number of disadvantages may result from this approach. For example, consider the following:

> (1) Does each employee have an equal opportunity to excel on these criteria measures: number of arrests, complaints, commendations. Number of arrests would be related to hours of the day worked and area patrolled. Number of complaints and commendations may also be dependent on factors not affected by job performance. (Consider the fairness of comparing commendations and complaints about a constable working the roughest section of town after midnight with the constable who is working with the Public Relations Section and is required to speak at junior high school assemblies.)
>
> (2) Are these measures representative of good job performance? A man with many days off sick in a given year may be an excellent policemen for the next 20 years.

The forced choice method of measuring job behavior is another means of obtaining supervisors' ratings. Police departments have seldom used this approach in obtaining a criterion. The rationale upon which this rating system is based is that of minimizing rater bias. The rater is required to select from equally favorable (or unfavorable) qualities for the person being rated. Only one of the equally socially desirable traits is positively related to job performance. (See Anastasi, 1968, p. 459, for an elaboration of this rating procedure.)

Dismissal vs. remain on the job as a criterion. This type of criterion compares the selection test scores of those who have remained on the job with those who have left. Those who have remained are considered successful job performers while those who have left are considered failures. There are certain relevant aspects of this criterion that should be considered.

(1) It does not take into consideration good and poor job performance of the present staff. Therefore selection tests validated in accordance with a criterion of this nature would not be predictive of good job performance but only of job tenure.

(2) Employees may leave the position for different reasons. A good employee may leave for a similar position in a more preferred location. His good test scores will be an inaccurate reflection of the failure (non-tenure) group.

(3) This type of criterion is best used when job training is very costly and on-the-job performance variability tends to be minimal. Although police training is costly, job performance skills differ widely. Tenure, therefore is important but should

not be considered the ultimate in importance. It should only serve as part of the criterion.

When considering terminators (within the experimental group) certain problems arise as to how they should be rated. Should terminators be give the lowest possible rating or should the reasons for termination be considered? Levy (1967) gave higher job performance scores to resigning employees over those who were requested to leave. The crucial question is: Should employees with good job performance records who resign serve within the success group, failure group, or an intermediate group? Should they be placed higher than those with a poorer job record but who are not resigning? As stated previously an important part of the answer to this question involves the expense of training the employee. If the expense is high, how much more economical will it be to keep a non-resigning low performance employee over a 20 year period than it is to risk training a proportion of good performing employees who then resign?

Training school marks as a criterion. The disadvantage of training school marks as a criterion is that their relationship to job performance is often unknown. Training school marks are usually a good measure of intelligence. The degree to which above average intelligence in a police training school class (which may be above the population averge for intelligence) relates to job performance is the crucial problem. Thweatt (1972) has reported that the IQ scores of the men who left the police department with which he was involved were higher than those who stayed. He stated: "Although one would want bright-normal men on the force, it is clear that the very bright do not stick with the force" (p. 67). In this situation extremely high training school marks were positively related to job termination.

Peer nomination as a criterion. This type of rating involves having fellow employees rate each other. The disadvantage of this approach is that it may result in unfair ratings. An employee may tend to under-rate good employees so as to improve his own chances of promotion.

Conclusions regarding criterion selection. It is suggested that the approach that appears most workable in gaining an accurate estimate of police job performance would be measures from supervisors' ratings and the method of paired comparisons. A combination of these two approachs could be used. It has been demonstrated through factor analysis (Ferguson, 1974) that the present supervisors' rating system can be represented by one factor a "Summary of Overall Performnce" measure. As one measure of job performance, sergeants could be required to estimate numerically a constables "Overall Performance". Secondly, because the main disadvantage of the paired comparison approach is the number of pairings that have to be considered, a rank order approach which would serve a similar purpose, to rank each man in comparison with his fellow workers, could be considered as another measure of job performance. Measures such as these should be highly representative of job performance.

Problems with Research of this Type: Validity of Criterion

Upon completion of this research a number of questions were presented regarding concepts and criteria of policing. A previous section has considered some of these questions by outlining the important qualities, roles, and concepts of "good" policemen. The acquiring of this information was very necessary because research variables related to qualities, roles, and concepts of policing had to be assessed at two different time periods. First during the selection of the recruit (independent measures); second, after time spent on the job (criterion measure). During the selection period a battery of psychological tests was given measuring candidates along the dimensions previously described. To measure job performance after job behavior has been observed presented another interesting methodological challenge. What is important in this area is the validity of the job performance measure. When the validity of a job performance measure is challenged, the question in its simplest terms becomes: Is this measure of job performance really measuring job performance? To consider this in some depth, the term "criterion" has been reviewed and clarified.

Criterion has been described as "a measure of the goodness of the worker" (Blum and Naylor, 1956, p. 174). Ghiselli and Brown (1948) have defined criterion as "any attribute or accomplishment of the worker that can be used as index of his serviceability or usefulness to the organization that employs him" (p. 62). Selecting the criterion that will be used to asses the performance of a policeman is a difficult problem. The validation of a battery of selection tests depends entirely upon the measures of employee success used. If the performance measures misrepresent the worker's job behavior then the selection tests which are chosen on the basis of a good predictive relationship with these measures will not be validated. For this reason and for many other usual reasons (consideration for promotion, merit awards, dismissal procedures, etc.) it is essential to have a sound criterion.

Certain concepts concerning a criterion have to be understood before an evaluative approach to the problem of the criterion can be made. A clear understanding of what comprises the ultimate criterion is essential. The ultimate criterion would be the perfect representation of how well a job was performed. This perfect measure would result in an assessment of every element of an employee's job behavior over the entire length of his stay in that position. This comprehensive assessment would have to include ratings from every person with whom the employee came in contact during his job, would have to take into account the amount and quality of the output of his work, the amount of training he required and the length of time he stayed with the position. It is, obviously, not possible to obtain an ultimate criterion. The ultimate criterion, furthermore, would be of little value in assessing the present employee because it could not be obtained until the employee had left his postion. For this reason distinctions are made amongst immediate, intermediate and ultimate criteria (Super and Crites, 1962). An immediate criterion generally emphasizes limited aspects of performance that can be acquired at the beginning of the job, for example, marks in training school. Unfortunately, measures such as these, may not be representative of later job performance. The intermediate criterion refers more closely to job performance (e.g., job performance ratings after one year on the job). Since the ultimate (perfect) criterion cannot be achieved, immediate or intermediate criteria have to be used instead. Because a less than perfect criterion has to be used a criterion deficiency will result (Blum and Naylor, 1956). This is the degree to which the ultimate criterion is not represented by the actual (immediate or intermediate) criterion.

For all job performance measures there is a criterion deficiency. The job performance measure is only an estimate of actual job behavior and therefore must fall short of the perfect or ultimate criterion.

In this research a criterion deficiency has also occurred. Part of this deficiency is based upon the solution to the problem: Whose concept of policing is being considered? It has been pointed out (Stirling, 1972) that various citizen groups have slightly different concepts of policing. In general, citizens desire policemen that are "helpful", "courteous", and "tactful". There concerns are directed primarily towards the policeman as he fulfills the public servant role. Under non-threatening conditions, citizens tend to place lower priority on the police roles of law enforcement and maintaining social order. It has also been shown (Pugh, 1975) that policemen themselves see the qualities and roles of the public servant as important. The policeman's first emphasis, however, appears to be the roles of maintaining order and law enforcement.

When, for this research, policemen were asked to judge the job performance of the subjects in this study, it was realized that their concepts of policing would shape their judgment. (Of course any rater's concept of policing would affect that rater's judgment of police performance.) It may even be possible that police sergeants' assessments may select the 'poorer' policeman (in terms of another societal grouping's definition) as 'very good' and vice-versa. This approach, then, has caused a type of criterion deficiency. Using policemen as raters will obviously result in the exclusion of citizens' viewpoints.

Justification for proceeding with this research. The method of using police sergeants as raters can be justified. It was mainly other policemen (staff-sergeants, sergeants and constables) that viewed the daily job performance of the subjects in this study. Also, police views on policing, are views based on experience and consideration of the wide spectrum of police work. For these reasons the use of police sergeants as raters was considerd. In any case, a survey of their views of policing was undertaken and was found to compare favorably with that of citizens in this community (Pugh, 1975). It was also pointed out that citizens generally favored police performance in this community (Alberta Bureau of Statistics and the Edmonton Police Commission, 1976). It was concluded that the police sergeants would give as valid a performance appraisal as was possible to obtain. It is recognized that this measure, as a criteron, has deficiencies that must always be considered when reviewing the results of this study.

It should be pointed out that other methods for obtaining a criterion are available. For example, it would have been interesting to have had raters assess policemen directly on certain qualities, roles or concepts (e.g.,courtesy, public relations skills) in addition to having them assessed on "overall performance". Whether the results would have been vastly different from what occurred in this study poses another research question.

What will follow now is the more formal review of the literature. In considering these studies the reader is cautioned to always take into account the job performance criterion that is being used.

CHAPTER II

LITERATURE REVIEW

Intellectual Level of Policemen and Police Recruits

Descriptive Studies: Average intelligence test scores of policemen and police recruits. Intelligence tests were among the first psychological tests to be developed. They were frequently used in measuring the intellectual development of police applicants and policemen. Primarily, for the earlier studies two measures of intelligence were used, the Standford Binet Intelligence Test (Terman, 1916) and the Army Alpha Intelligence Test (War Department, 1919). Many of the studies that were conducted before 1940 are reported in Table 3. As can be seen from the average percentile scores most police departments in these earlier times had policemen with above average intelligence.

Many of these earlier studies concerning the intelligence of police officers were gathered from the "National Crime Commission's Report on the Abstract Intelligence of Police Officers" (1928). This early report recommended that a minimum intelligence score of 65 on the Army Alpha (57th percentile) would be a good standard when considering police recruits, with a score of 105 (85th percentile) as a desirable additional qualification. For

Table 3

Early Intelligence Scores
of Police Applicants and Policemen

Group	n	Standford Binet IQ	Percentile
Policemen/Firemen Applicants, Detroit Terman and Otis, (1917)	30	median = 80	26
Police Applicants Detroit Ostrander (1925)	108	median=13.1 yr.	40
		Army Alpha Score	
Police Officers Detroit, Thurstone (1923)	17	Lieutenants, mean = 58	50
	34	Sergeants, mean = 55	47
	307	Patrolmen, mean = 71.4	64
Applicants for patrolmen, San Diego (1928)*	64	median = 76.5	67
Men in Maryland State Police Training School (1928)*	50	median = 100	82
Men in police school of instruction Los Angeles, (1928)*	321	median = 116	86
Policewomen Washington, D.C. (1928)*	155	median = 144	96
Policemen Berkley, California (1928)*	26	median = 149	97
Applicants in Palo Alto, California Merril (1927)	113	mean = 104	82

*as reported from the National Crime Commision (1928) police sergeants an Army Alpha score of 90 (68th percentile) was seen as a minimum requirement.

Problems were noted in the National Crime Commission report when the intelligence of a policeman was extremely high. Two police departments (Palo Alto, California and Berkley, California, see Table 6) indicated that those policemen with very high intelligence were leaving police work in a relatively high frequency. The following conclusions were reported.

> It will be noted ... that the men who voluntarily left the Palo Alto police force made the astonishingly high mean Army Alpha score of 171.5 (98th percentile). Mr. August Vollmer reports too that in Berkley he had some difficulty with police officers whose Army Alpha score is unusually high. It would be a fallacious procedure to draw positive conclusions from such meager data but the Palo Alto facts and Mr. Vollmer's observations may be indicative of the fact that there is a maximum amount of abstract intelligence which may eliminate applicants for police work as certainly as failure to possess a minimum amount. (p. 78)

These early studies have shown that the intelligence of police officers varies from department to department but generally is above average. Also evident is the possibility that extremely intelligent police officers tend to leave the police force at a higher frequency then others.

A different conclusion has been draw from another study. Bain (1939) in an often cited article stated that "at least 75 per cent of the policemen in the country are mentally unfit for their work" (p. 452). This statement was based on the assumption that policemen should score above 120 on the Army Alpha intelligence test. This would require policemen to be at the 90th percentile or above in intellectual functioning. In the light of previous findings up to this time this statement calls for very high intellegence test scores with no validating information to insure that those with very high scores would be the best policemen. In fact, as has been previously cited, it appears that highly intelligent policemen during this time period tended to leave police work more often than others. Bain's conclusions have to be considered in terms of the times to which they referred. It appears that he was calling for more intelligent policemen during a time period when other studies had indicated that the highly intelligent policemen did not remain with the job. Because the role of the policeman is changing it is suggested that in 1939 Bain's conclusion may have been in error. In more recent years, however, the very intelligent policemen have remained on the job as will be pointed out below.

More recent studies involving intelligence tests and police recruitment also have indicated that intelligence test scores were generally above average but varied widely. A number of studies were based on New York City policemen. Neiderhoffer (1967) reported that: "During the past ten years the average IQ for a class of recruits at the New York City police academy has been approximately 105" (p. 33). McManus, Griffin, Wetteroth, Boland, and Hines (1970) have pointed out that the average IQ of New York City recruits varied considerably. For example they reported that recruit test results on the Otis Quick Scoring Mental Abilities Test (standard error of measurement = 3.0) (Otis, 1939) for the graduating class of October 1964 averaged 109.55 (75th percentile) while in May 1965 the average was 93.19 (32nd percentile). The authors concluded that the reasons for the high variability of intelligence test scores within their research was dependent partially on the availability of potential recruits. It was suggested that standards for the acceptance of recruits changed according to external factors (health of economy, unemployment, educational possibilites, salaries, etc.). These external factors appear to be directly related to the year to year variation in the average intelligence scores of new policemen.

Fenster and Locke (1973) examined the intelligence of experienced New York policemen with an average age of 31.51 years. The authors compared 178 policemen enrolled in college courses with 174 policemen who were never enrolled in any college courses. Two matched control groups were also tested; 93 college civilians and 178 non-college civilians. The average scores of these four groups on the Otis Quick Scoring Mental Abilities Test were: college enrolled policemen, 114.79 (82nd percentile); non-college policemen 110 (75th percentile), college civilians 116 (83rd percentile), non-college civilians 102 (51st percentile). These authors concluded: "It can safely be said that the intelligence of the average policeman in New York City is significantly higher than that of the general population" (p. 280). This conclusion when considered after reviewing the other New York studies appears to be correct.

In Portland, Oregon a number of descriptive studies of the average IQ of policemen have also been conducted. Kole (1962) and Matarazzo, Allen, Saslow, and Weins (1964) reporting from the same data found that average Wechsler Adult Intelligence Scale (WAIS) IQ score for a group of successful police applicants in Portland, Oregon was 112 (80th percentile). The authors commented on this high average by stating the following.

> Thus, as a group, they represent men who, endowed with good intellectual abilities which they have not fully developed or tested, often in their own words turn to police work and firefighting as career choices next best to the business and other professions for which they qualify but for which their educational limitations, lack of knowledge, or lack of opportunity prevent them from entering. (Matarazzo, et al., 1964, p. 133)

Two other studies based on this same police department in Portland Oregon also have reported above average intelligence scores for police officers. Matarazzo, Wiens, Matarazzo, and Manaugh (1973) found an average IQ score on the WAIS of 118.1 (89th percentile) for 29 police recruits. Zaice (1962) compared the Army General Classification Test (AGCT) (War Department, 1947) scores of patrolmen, detectives, and command personnel in the Portland Oregon police department and also found above average scores. He obtained the following results: patrolmen 123, detectives 126, command 130. Zaice's results suggested that the more intelligent policemen were remaining with the Portland police force and being promoted. The overall average for Zaice's three groups placed these Portland policemen at the 90th percentile. All of the studies reporting on subjects from the Portland, Oregon police department have indicated a very high average intelligence for its policemen for this recent time period.

It should be noted that the Portland Oregon police department may not be typical of most police departments. Fenster and Locke's (1973) study, which was reviewed previously, pointed out the following.

> However, it should be noted that ... 40% of the Portland, Oregon, police department is college educated; this is not typical throughout the country. (p. 276)

This fact should be considered when reviewing studies based upon members of the police department in Portland. It would be in error to develop conclusions about policemen in general after reviewing studies based only on policemen in Portland.

Criterion Studies: The relating of intelligence test scores to job performance measures. Of high importance for police personnel departments are studies which relate police performance with intelligence test score results. Thweatt (1972) compared the intelligence scores of new policemen who left the police force with a norm group of remaining experienced policemen. He found that the group mean for the intelligence scores (Otis Test of Mental Ability) of fifty experienced officers and sergeants on the Tuscon Arizona police force was 111 (76th percentile). From a group of new recruits he reported that the IQ scores of those who dropped out of the force was 118 (89th percentile) and that this mean score was significantly higher than the norm group of the fifty policemen. He stated:

> Although one would want bright normal men on the force, it is clear that the very bright do not stick with the force. (p. 62)

These findings tend to support those reported in an earlier study (National Crime Commission, 1928). Thweatt has concluded that the "bright normal" men did not remain with police work as often as those in the average range of intellectual development. Thweatt suggested that these results implied that policemen in the 75 to 100 percentile range for intelligence test scores may have a higher dropout rate then other policemen. This conclusion, however, was not well substantiated. Thweatt's study failed to indicate the IQ scores of the recruits that did not resign. Only the comparison in IQ between the dropouts and the experienced norm group was reported. This omission weakens the impact of the conclusions that can be drawn from Thweatt's results.

Dubois and Watson (1950) compared the intelligence test scores of the two classes of police recruits (n =129) in St.

Louis, Missouri with a number of follow-up variables which included marks in training school and performance ratings after ten weeks on the job. Dubois and Watson found that the Army General Classification Test (AGCT) correlated with training school marks for the two classes, r=+.54, r=+.50. This result suggests that training school marks can be partially estimated from intelligence test scores. When the AGCT was correlated with performance ratings, the correlation was low, r=+.10. This on-the-job performance rating correlation can be questioned. These new policemen were rated after ten weeks of work as policemen. It would be difficult for a beginning policeman to demonstrate his proficiency after this short period of time. Also difficult, would be rating an individual after observing him at his job for only the first ten weeks.

Mullineaux (1955) gave this same test, the AGCT, to recruits in the Baltimore, Maryland police department. He reported that the AGCT correlated with police academy marks, r=+.73. This finding supports those of Dubois and Watson (1950) indicating a positive relationship between intelligence test scores and marks in the training academy. Also in support of this finding is the study of King, Norrell, and Erlandson (1959). They have reported that a language test score and a reading test score yielded the best prediction of first term grades for policemen in the police academy, R=+.56. The studies of Dubois and Watson, 1954, Mullineaux, 1955, and King et al., 1959; all are in agreement that marks in police training school can be predicted from intelligence test score results.

Similar results were also found in the study of Mills, McDevitt, and Tonkin (1966). They tested two groups of Cincinnati Ohio recruit policemen. The AGCT scores correlated with rank in training class, r=+.595, r=+.780. The AGCT scores, however, did not differentiate between those who completed training and those who did not. From these results it appears that little or no relationship between dropouts and intelligence scores exist. This finding does not support Thweatt's (1972) conclusion that "the very bright do not stick with the force" (p. 62). (It should be noted that Thweatt's conclusions were drawn from post academy performance measures and therefore are not entirely comparble to the study of Mills, et al. 1966). In general however, a number of studies have suggested that intelligence test scores and police training marks are highly related.

Bass, Karstendiek, McCullaugh, and Pruitt (1954) correlated the Wonderlic Personnel Test (Wonderlic, 1961) with performance ratings and found correlations of r=+.19 for 37 city police in Baton Rouge, Louisiana, and r =+.26 for 22 Baton Rouge area deputy sheriffs. These results comparing intelligence test scores with job performance ratings are similar to those of Dubois and Watson (1950). The correlation between intelligence test scores and performance ratings appears minimal, being slightly positive.

From these studies concerning the intelligence scores of police officers and the relationship of these scores to job performance measures the following conclusions can be drawn.

> (1) The average intelligence of policemen varies among police departments. In the majority of articles reviewed the intelligence of policemen was above the population average.
>
> (2) Intelligence test scores of police recruits appear to be significantly correlated with their marks in police training.
>
> (3) Intelligence test scores are not highly correlated with performance scores. This relationship appears unclear. It has been reported that extremely intelligent policemen tend to leave police positions more often then other policemen. This occurrence could cause a curvilinear relationship between intelligence test scores and performance ratings.

Interest and Aptitudes of Police Groups

Early testing to discover police interests and aptitudes began in the 1930's. Diehl, Patterson, Dorvak and Longstaff (1933) found that a superior group of policemen (captain's ratings) did better on the Pressey Clerical Aptitude Test than did the average and below average group. The authors stated:

> Such tests would seem to have little value in identifying the highly efficient policemen; their only value seems to be that they identify the inefficient policemen. (p. 45)

With reference to mechanical ability the authors stated that "... policemen are not different from men in general in their ability to mechanically manipulate" (p. 53). They also found that as a group policemen were superior to the general population in clerical speed and accuracy. The authors reported that the "superior" police group did much better in number and name checking than did the "poor" group. From these results it appears that superior policemen in the 1930's scored well in comparison with populations norms on tests indicating clerical skill which included number and name checking.

Results from the Strong Vocational Interest Blank (SVIB) have often been reported for policemen. It should be pointed out, when interpreting SVIB results, that the criterion groups with which the test-takers are compared are divided as follows for each scale of the test. Three fourths of the men of the criterion group (e.g. policemen choosen to establish the norms for the "policeman" scale) score in the 'A' category, one fourth in the 'B' category, "a rating of 'C' means 98 to 100 percent of the criterion group score higher" (Berman, Darley, and Patterson, 1934, p. 217). These categorizations should be referred to when considering the results of the SVIB.

Berman, Darley, and Paterson (1934) were among the first to examine the interests of policemen as measured by the Strong Vocational Interest Blank (1933). They used test results of policemen from the city of Duluth. This group of policemen formed part of Strong's original criterion group for the "policeman" scale. This scale was developed to separate interests of policemen from the interests of "men in general". The experimental procedure included having the captain rate the policemen for performance efficiency. The ratings resulted in the following groupings: Group I--43 subjects above average, Group II--44 subjects average, Group III--36 subjects below average. On the "policeman" scale all three performance groups of policemen averaged in the 'A' category of "policeman" interest indicating that high police interest (as compared with "men in general") was not correlated with good job performance. The authors reported interest scales that differentiated the policemen according to job performance.

> Groups I and II are in every case more interested than Group III in occupations involving social contacts (YMCA Physical Director, Personnel Manager), whereas the lowest efficiency group tends to be undifferentiated from people in general with respect to such interest. (p. 228)

From this early study two conclusions are suggested. First, the "policeman" interest scale was valid for predicting police interest but not valid for predicting police performance. Second, above-average policemen showed more interests in social contacts than did other policemen.

These results suggest that because policemen do not differ greatly on the "policeman" interest scale, interest scales other that the "policeman" scale may be better predictors of future job performance.

Fraser (1949) had the Strong Vocational Interest Blank sent to 325 R.C.M.P. who were centered around Edmonton Alberta. Three humdred and eight subjects completed and returned the answer sheet. These volunteers had been previously divided into four groups, three criterion groups based on performance efficiency ("most efficient", "mediocre", "least efficient") and a fourth group of new members. All four groups averaged in the 'A' category on the "policeman" scale. It was also shown that the "efficient" policemen scored significantly higher on "personnel director", "public administrator", and "occupational level" and significantly lower on the scales "musician", "architect", "dentist", "engineer" and "chemist". Fraser's results suggested that policemen have high interests in directing others and low interests in many positions that could be described as "professions".

Kates (1950) tested 25 volunteer New York City patrolmen with the SVIB, a job satisfaction questionnaire, and the Rorschach (scored with the Munroe Inspection List which includes a measure of maladjustment). The mean "policeman" scale score was 40.6 (category 'B'). This score is significantly lower than the criterion group of policemen used by Strong. Kates's measure of job satisfaction correlated with the "policeman" scale, r =+.35. This correlation was in the positive direction and, although small, indicated that a significant relationship between job satisfaction and job interest as measured by the SVIB occurred.

Dubois and Watson (1954) also used the SVIB in testing two recruit classes. They found that the scores of most of the recruits on the "policeman" scale were in the 'A' or 'B' category. Academic grades in the police academy were slightly negatively correlated (r = -.09, -.12) with the "policeman" scale. These correlations were insignificant and may have been caused by a chance effect. However, the inverse relationship between police academy grades and police interest may indicate that among a group with above average intelligence the slightly lower in intelligence (although still above the population average) have interests similar to the criterion group that was used to develop the "policeman" scale.

The authors also gave a speed perceptual task (Figure Matching Test) which correlated significantly with academy grades, r = +.29. They also found that other object manipulation tests (Bennett Mechanical Comprehension Test and the Minnesota Paper Form Board Test) correlated positively, r = +.28, +.29, +.38, +.29) with grades in the police academy. A measure of job performance after 10 weeks on the job showed no significant correlations with any of the variables in the test battery. This finding may not have been unexpected as 10 weeks of job performance is a limited time period over which little differentiation among recruits would be expected to occur and be noticed by raters.

Zaice (1962) reported a descriptive study on 104 experienced policemen in Oregon. The scores on the "policeman" scale of the SVIB ranged from 22 to 62 with a mean of 45 (which falls within the 'A' category). When policemen were separated according to rank their mean scores on the "policeman" scale were: command, 43.4 ('B+'); detective, 46.0 ('A'); patrolmen, 45.3 ('A'). It should be noted that the command group had an 'A' average on "public administrator". The other noted high scores for the entire group were: "production manager", "farmer", "printer", "aviator", "public administrator", "senior CPA", and "credit manager". This study was not predictive of future performance as a policeman. It offered no correlational coefficients between interest scores and job performance measures. The results, however, did serve to validate the "policeman" scale. The entire police group, as mentioned, averaged within the 'A' category on the "policeman" scale.

Matarazzo, Allen, Saslow, and Wiens (1964) gave the SVIB to policemen (n=112) and firemen (n=124) in Portland, Oregon. Police applicants scored highly ('B+' or more) on:

"farmer", "aviator", "printer", "math-science teacher", "policeman" (average = 46, or 'A'), "public administrator", "senior CPA". When comparing policemen and firemen applicants differences were noted.

> Young policemen are more oriented towards jobs involving working with people (...), while young firemen are oriented towards occupations requiring work with one's hands or the business world.... It was our clinical impression that firemen and policemen applicants do differ with the former being the rugged, outdoor, family handyman type of person and the latter the more intellectual, professional-type person. (p. 129)

This research again validated the "policeman" interest scale by differentiating this police group from the norm group of men-in-general. No predictive relationships between interests and job performance were reported.

Thweatt (1972) tested 50 experienced officers and sergeants on the Tuscon Arizona Police Force with the SVIB. He also tested an experimental group of 105 recruits. The experienced group had a mean on the "policeman" scale of 34.86 ('B-'). Thweatt reported the following concerning the "policeman" scale score.

> This statistic is somewhat misleading as three distinct groups tended to show up. About one third of the men had scores highly like policemen; about one third had scores in the social service areas, i.e., like social workers, and the last third fell into what among police applicants might be called "frustrated professional scores", i.e., doctors, lawyers, engineers, etc. (p. 61)

Among the recruits Thweatt reported that the scores of the dropouts were significantly lower on the SVIB "policeman" scale than were those of the norm group. "Among the dropouts, all the patterns fell into the category of 'frustrated professionals' previsouly described as unresolved envy for prestige occupations" (p. 61). (It should be noted, as mentioned previously, that Fraser's results (1949) also indicated that the most efficient policemen had low interest in other professions). In Thweatt's study the dropout group had a mean of 25.76 on the SVIB "policeman" scale compared with the norm group mean of 34.86. The mean score of the remain group of recruits is not given. If Thweatt's results were erroneously generalized to the entire police population it might be concluded that the "policeman" scale is losing its validity for predicting interest in police work. A score in the '-B' category for the norm group of policemen on the "policeman" scale suggests that the scale was not an accurate reflecton of police interest for the policemen in Tuscon involved in Thweatt's study.

If all of the studies previously cited concerning the SVIB are considered, it may be concluded that the "policeman" scale has had fair success in discriminating policemen from men-in-general. The averages of this scale on studies reported above are as follows: Berman, et al. (1934), scale score of 45 ('A'); Fraser (1949), scale score over 45 ('A'); Kates (1950), scale score of 40.6 ('B'); Dubois and Watson (1954), scale score over 35, ('A' or 'B'); Zaice (1962), scale score of 45 ('A'); Thweatt (1972), scale score of 34.86, ('B-'). For the purpose of prediction, however, none of these studies has reported any significant correlation between the "policeman" scale and on-the-job performance measures.

In a recent article by Johansson and Flint (1973) the up-dating of this "policeman" scale to present population norms was reported. The criterion group selected were 196 Minneapolis Minnesota policemen who "had at least three years experience and ... liked their job" (p. 40). The responses of these policemen were contrasted with the responses of a men-in-general sample of 1000. Seventy-four items that differentiated the two groups were aggregated to form a new "policeman" scale. The authors reported the following characteristics of the policeman group.

> The items the policemen "liked" more frequently than the men-in-general group fell into three broad categories: militaristic, risky, and mechanical. For example, a greater percentage of policemen than men-in-general said they would like to be a "military officer", participate in "military drill", be with "military men", and "drill soldiers". They responded more favorably to risky types of activities, such as being an "auto racer", an "airplane pilot", or a "secret service man", "pursuing bandits in a sheriff's posse", and engaging in "thrilling dangerous activities over quiet safe activities". The policemen also felt they had "mechanical ingenuity" and expressed more interest in mechanical avocations than the men-in-general group by making positive responses to such items as "auto mechanic", "carpenter", "shop foreman", "mechanical drawing", and "adjusting a carburetor".
>
> ... Results showed that the Minneapolis policemen

> had about the same social service types of interests as the men-in-general sample, but two items differentiated the groups---policemen had a greater liking for "giving first-aid assistance" and considered themselves "able to smooth out tangles and disagreements between people". (p. 41)

This scale was validated on a number of police groups and the results are reported in Table 4.

One interesting finding from this research is that on the masculinity-femininity scale policemen had a very high mean, 59, "indicative of strong authoritarian-structured types of interests" (p. 42). Because of the importance of authoritarian attitudes the relationship of this variable with job performance should be considered in future selection studies of policemen.

In contrast to the previous study that found differences between policemen and the average citizen, the study of Thorndike and Hagan (1959) found no differences between policemen and a selected group. They reported interesting results upon following the careers of 17,000 airforce cadets tested for intellectual, numerical, perceptual-spatial, mechanical and psychomotor aptitudes during World War II. One hundred and nineteen of these men had become policemen by 1955. It was reported that the policemen did not demonstrate any very specialized aptitudes as measured by the original tests. Except for eye-hand coordination the police group scores were below the average airforce cadet but within the normal range for this group.

Table 4

Means Scores on the "Policeman Scale"
As Reported By Johansson and Flint (1973)

Sample	N	Mean	S.D.
1969 Minneapolis policemen	96	50.0	10.0
1969 recruits (tested 12/69)	25	41.2	9.8
1969 recruits (tested 6/70)	22	45.0	11.7
1969 recruits (tested 9/71)	21	46.3	9.1
1970 recruits (tested 4/70)	25	48.3	8.2
1970 recruits (tested 9/71)	23	44.8	9.0
1971 recruits (tested 9/71)	29	47.4	10.4
1967 Utah County Sheriffs	111	50.4	9.4
1968 Minnessota Highway Patrolmen	288	49.4	9.2
1968 Racine, Wisconsin Policemen	44	52.1	9.7
1968 University of Missouri Policemen	38	47.0	10.5

In recent years the approach of correlating aptitudes with job performance has been used infrequently. Holmes (1951) tested 56 civilian security guards in Benicia Arsenal, California. He found that technical reading (patterned after U.S. Army technical bulletins) and a spatial relations test produced a combined R of +.50 when correlated with a composite rating of success on the job. The author reported that the subjects score on 'spatial relations' correlated with success, r = +.43. He commented as follows.

> Its (spatial relations) importance could depend upon the value of such an ability in predicting the discriminating acuity of a patrolman in detecting whether or not the physical objects on his beat had or had not been tampered with. (p. 399)

Marsh (1962) also correlated aptitudes with the job performance of 619 deputy sheriffs. He reported that a clerical task, number series completion, was the best predictor of job performance after 8 years on the job. It was indicated that those who scored above the 73rd percentile were more likely to be successful.

Azen, Snibbe, and Montgomery (1973) followed up Marsh's (1962) study and reported furthur findings for a 20 year period. They concluded that:

> The principal result of the stepwise-discriminant analysis is that the Kuder Mechanical score emerges as the most generally useful predictor of the criteria (since it predicts 3 of the 6 criteria). (p. 191)

The Kuder Mechanical score was positively related to "job type" (patrol or other), "average supervisor's ratings" and "rank status" (promoted or not). A discriminant analysis probability of correct classification ranging from 63% to 72% was demonstrated. These results indicate that mechanical ability as measured by the Kuder Vocational Preference Record was a good predictor over a 20 year period of job performance. This interest in "mechanical" gives support to the previously reported findings of Johansson and

Flint (1973).

Blum (1964) gave a battery of tests to each recruit hired in a "major metropolitan police force" in the years 1956-57. Seven years later each recruit's file was reviewed and performance measures on a variety of criteria were estimated. The SVIB was one test that was included in the battery and the following results were reported. Number of vehicle accidents correlated postively with: "physician" scale, r=+.30; "psychologist" scale, r=+.28; and "physicist" scale, r=+.26. Number of serious formal charges correlated r=+.23 with the "author-journalist" scale. Serious misconduct correlated r=-.28 with the "physician" scale. The "carpenter" scale correlated r=-.24 with commendations. The "policeman" scale correlated r=+.22 with days lost due to illness. Injuries correlated r=+.22 with the "physicist" scale and r=-.18 with the "carpenter" scale. Blum has stated that: "Rather oddly the carpenter interest scale most often correlated, always negatively, with performance" (p. 130).

The unexpected (but low) positive correlations between the professions of "physician", "psychologist", "author-journalist", and some of the negative performance measures supports Thweatt's hypothesis that within a police population there is a group of "frustrated professionals". Blum's results tend to validate this conclusion.

A number of other studies have included measures of police interests. These studies, along with those already presented are shown in Table 5. The studies are grouped according to frequently occurring categories. Those variables that are correlated with a criterion are indicated by 'CWC' (Correlated With Criterion).

The interest areas that seem to occur most frequently related to good job performance are the mechanical ("mechanical ability", "technical reading", "spatial relations") and clerical ("number and name checking", speed perceptual tasks). High scores on scales such as "physician", "psychologist", "physicist", "author-journalist", and "carpenter" appear to have an inverse relationship to job performance. A replication of these relationships with later on-the-job performance would provide police selection officers valuable information for interpreting police interest test scores.

Personality Characteristics of Policemen

Personality characteristics are often cited as being very important when considering applicants for the postion of policemen. Vollmer (1947) recognized this problem.

> All individuals have certain personality traits and characteristics which are not observable in routine police examinations and interviews, and which can only be detected by experts trained in this work. Unfortunately, many of these characteristics will render an individual completely unfit for police service, regardless of his other qualifications. Among those traits are inability to get along with fellow workers,

Table 5

Interest Areas of Policemen

AUTHOR AND YEAR	INTERESTS AREAS NOTED
SOCIAL CONTACTS	
Berman, Darley, and Patterson(1934)	Above average and average police groups were more interested in social contacts than the below average group. They were high on 'YMCA director' and 'personnel manager'. (CWC)
Spaulding(1948)	Police applicants on Kuder high on 'social service'.
Gough(1955)	Police score over B+ on 'math/science teacher'.
Matarrazzo, Allen, Saslow and Wiens(1964)	Police applicants low on 'minister'.
Thweatt(1972)	About one third of experienced officers had high scores in social service area, like social workers.
SPORTS	
Thorndike(1959)	Policemen score above average on eye-hand coordination. Biographically information indicated they did well in physical training.
Dillman(1964)	High rated group of policemen mention sports as the nicest experience in youth.
MEMORY AND OBSERVATION	
Holmes(1951)	Good security gaurds score higher on spatial relations (CWC, r= +.43)
Germann(1958)	Cites study where memory and observation found to be unimportant.
Super and Crites(1962)	Police average around the 27th percentile on spatial visualization.
NUMBER AND NAME CHECKING	
Diehl and Patterson(1933)	Superior group better in number and name checking than poor group. (CWC)
Holmes(1951)	Good security gaurds high on technical reading (patterned after U. S. army technical bulletins). CWC, r = +.50.
Marsh(1962)	Found that police applicants who scored over the 73rd percentile on number series completion were more apt to be successful.
Dubois and Watson(1954)	Silhouette matching correlates with final marks, r= +.25,+.29.
MECHANICAL ABILITY	
Diehl, Patterson, Dorvak, and Longstaff(1933)	Policemen are not different from men in general on the ability to "manipulate mechanically".
Diehl and Patterson(1933)	Superior police scored higher on O'Connor Tweezer Dexterity Test than the below average police group. (CWC)
Dubois and Watson(1954)	The Bennett Mechanical Comprehension Test correlated with final marks, r= +.28,+.29.
CLERICAL TESTS	
Diehl, Patterson, Dorvak and Longstaff(1933)	A superior group of policemen did better on a clerical aptitude test than did the average group. (CWC)
Spaulding(1948)	Police as a group score low on 'computational' and 'clerical'.
Gough(1955)	Policemen score high on 'printer'.
Matarrazzo, Allen, Saslow and Wiens(1964)	Policemen score high on 'printer'.
POLICE INTEREST SCALE ON THE STRONG VOCATIONAL INTEREST BLANK	
Berman, Darley and Patterson(1934)	The policeman scale did not differentiate superior and poor policemen. (CWC)
Kates(1950)	The mean score on this scale for a group of experienced policemen was B+ (40.6).
Dubois and Watson(1954)	Scores of most recruits were A or B+. There was a slight negative correlation with grade.
Gough(1955)	Policemen score over B+.
Zaice(1962)	Tested three groups of policemen; scores on this scale were: command 43 (B+), detectives 46 (A), patrolmen 45 (A).
Matarrazzo, Allen, Saslow and Wiens(1964)	Police applicants group average A.
Thweatt(1972)	Experienced policemen average 34.86 (B). Scores of terminators from force significantly lower, 25.76. (CWC)
LEADERSHIP	
Gough(1955)	Policemen average B+ on ' public administrator'. They also score over B+ on 'interest maturity'.
Zaice (1962)	Policemen score high on 'senior CPA','production manager', public administrator'.
Matarrazzo, Allen, Saslow, and Wiens(1964)	Police recruits score high on 'public adminsitrator' and 'senior CPA'.
ADVENTURE	
Gough(1955)	Policemen high in 'masculinity' and 'aviator'.
Matarrazzo, Allen, Saslow and Wiens(1964)	Police applicants high on 'aviator'.
Zaice(1962)	Policeman high on 'aviator'.
FARMER	
Gough(1955)	Policemen score high on 'farmer'.
Zaice(1962)	Policemen score high on 'farmer'.
Matarrazzo, Allen, Saslow and Wiens(1964)	Police applicants score high on 'farmer'.
AREAS OF LOW INTEREST FOR POLICE	
Kates(1950)	With experienced policemen 'occupational level' correlates negatively with job satisfaction.
Thorndike(1959)	Biographical information on police show them with limited education, few books at home, having done poorly in algebra, physics, and trigonometry.
Matarrazzo, Allen, Saslow and Wiens(1964)	Police applicants low in: 'artist', 'psychologist','architect', mathematician','physist', 'advertising', 'lawyer', 'author-journalist'.
Thweatt(1972)	He describes interest scores on a police recruit group One third were "frustrated professionals". Most of those from this study that left the force were in the "frustrated " group.

> refusal to accept orders and instruction, cowardice, tendency toward brutality, and a sense of superiority which jeopardizes the probability of good public relations. Such persons if appointed, will inevitably have a detrimental effect upon the department. Frequently these characteristics will not manifest themselves until the officer has passed his probationary period, and the problem of removing him from the force becomes difficult (p. 189).

Although Vollmer realized at that time that no tests were available that could "... infallibly measure every capacity of the human personality" he believed that a scientific approach to the problem should be developed. Studies have shown that ten percent of the population have or will have emotional problems. Police departments must be certain that their selection procedures will screen out individuals applying from this segment of the population.

Roe (1956) has supported the concept of a relationship between personality and choice of work.

> No truly comprehensive work has been done with personality tests as such in the field of occupational psychology. There are many studies of particular groups by personality inventories, and a few with projective and other techniques. Although the evidence is not extensive, there nevertheless seems to be no doubt that some specialized occupations, at least, do attract persons who resemble each other in some personality characteristics (p. 80).

A number of studies have appeared in the literature dealing with the personality traits of police applicants and experienced police officers. The studies reviewed will be grouped into three categories: descriptive studies that compare policemen's scores with test norms, descriptive studies that compare policemen's scores with the test scores of other groups, job related studies that compare policemen's scores with present or future job performances.

Descriptive studies comparing policemen with test norms. These studies involve administering to a group of policemen a personality test and then comparing the average scores obtained by the police group with the average scores of the general population as supplied by the test designer. The main problem with this approach is the type of knowledge that is obtained. Knowing the averages of a police group and a normative group on personality test variables and where individual police applicants score in relation to these averages gives little or no substantial information about the policeman's future job performance. These test measures have to be correlated with job performance measures to have predictive value. However, test variables which indicate policemen are different from the average can be the beginning points for hypotheses concerning test variables that will be predictive of future job performace criteria.

Rappaport (1945) gave the Rorschach to a group of 54 randomly chosen Kansas State Highway Patrol. He reported that the police group were not representative of the general population. They were withdrawn, lacked colorfulness, and showed a limited range of interests. He commented on the general personality characteristics of this group of Kansas

State Highway Patrolmen.

> ... (A) great percentage of these patrolmen came from farm environments, and if compared with a city population would have to be considered in some degree schizoid (p. 29).

As Rappaport has pointed out the rural nature of this group of policemen may be responsible for the qualities he has noted. The author's comments should be considered when examining candidates from a rural background. It should not be forgotten that these results may have less relevance at this time due to the vast changes that have occurred in the past thirty years.

Kates (1950) also used the Rorschach in a study involving twenty-five New York city patrolmen who volunteered to take part in his research. He used the Munroe Inspection List which gives a score of maladjustment on this test.

> The degree of maladjustment of the policemen, as measured by the Rorshach Test was slightly but insignificantly greater than that of biologists. Probably, the policemen, as a whole demonstrated as many signs of maladjustment as may be found in other groups (p. 251).

These results indicate that this volunteer group demonstrated an average level of maladjustment. Kates also compared the maladjustment scores with a measure of job satisfaction. He found that job satisfaction correlated positively, r=+.47, with the maladjustment score. This means that the higher the policeman's maladjustment score the more it indicated job satisfaction. This correlation may reflect a trend in policemen to develop a more "maladjusted" style of interacting with the public. The small size of this group and the fact that these subjects were only those who would volunteer their free time to take part in the research suggests that the results are subject to question and the issues involved need furthur study before conclusions can be drawn.

Sterne (1960) gave the Kuder Preference Record (Personal) to the majority (n=49) of a small municipal force. He found that this group scored significantly higher than the test norms on the "directing others" scale and significantly lower on the "seeking to avoid conflict" scale. Policemen as the author predicted have personality qualities that enable them to lead others and face conflicting situations.

Three authors have tested police groups with the use of the Edwards Personal Preference Schedule (EPPS) and compared their results with the test norms. Zaice (1962) gave the EPPS to a "representative group of patrolmen, detectives and command" personnel from the Portland, Organ Police Department. Matarazzo, Allen, Saslow and Wiens (1964) gave the EPPS to 243 successive civil service applicants for policemen and firemen also in Portland, Oregan as part of the application procedure. Simon, Wilde and Cristal (1973) used the EPPS to test 38 male members of a police force in a county of New York State enrolled in a college course. The results of these tests are shown in Table 6 and indicate the test scales that discriminate police groups from population norms.

In general the studies are in agreement. Zaice's study which tested the most representative sample of policemen had confirmation on all of the significant findings he reported by at least one other study. He found that the average policeman had high scores (above the 60th percentile) on the scales of "achievement", "exhibition", "heterosexuality", and low scores (below the 40th percentile) on "affiliation", "nurturance", and "abasement". Zaice commented on the type of person with these high and low trait characteristics.

> He does his best to be successful and attempts to accomplish tasks requiring skill and effort. He needs to be a recognized authority and will normally attempt to do difficult tasks well. He cannot be said to be shy for he seeks attention and is known to tell amusing stories, witty and clever sayings.
>
> ... The policeman is not the type to show a great deal of affection. His job is such that he must recognize violators of the law and impartially take necessary steps to regulate members of of society, according to the mandates of the law.
>
> ... They are willing to be part of an organization and take orders (p. 58,59).

From these descriptive studies that compare policemen with test norms the qualities of striving leadership ("achievement" and "directing others"), courage ("not avoiding conflict"), an average level of adjustment, and a

Table 6

Police Scores on the Edwards
Personal Preference Schedule

	Zaice (1962)	Matarazzo, et al., (1964)	Simon, et al., (1973)
Policemen higher than population norms.	achievement exhibitionism heterosexuality	achievement exhibitionism heterosexuality intraception endurance*	achievement exhibitionism hetereosexuality aggression*
Policemen lower than population norms.	nurturance affiliation abasement	nurturance aggression* autonomy succorance	nurturance affiliation abasement order endurance*

*Opposite results reported in one of the other studies in this table.

low need for affection and sympathy from others (low scores on "abasement", "affiliation" and "nurturance") appear to be the predominant qualities of the average policeman when compared with general population norms.

Descriptive studies that compare policemen with other groups. A number of authors have compared policemen with other groups and have pointed out differences. These studies give general characteristics of the average policeman in comparison with other groups. One question to consider with these studies (and also relevant to the group of studies just previously discussed) is their validity.

Neiderhoffer (1967) has commented upon the policeman's "glass house" approach to studies by social scientists. Policemen, he states, fear "learning too much about themselves." It appears only reasonable that policemen when asked to take part in these descriptive personaltiy studies would try to present the best image of themselves possible. The differences that occur in comparison studies between police groups and other groups may be a result of the socially desirable personality that the members of these groups wish to portray. The results of these studies should be considered with this possible interpretation of the policeman's approach to research of this type. It is possible that many studies presenting police personality profiles should be considered as an aggregrate role conception that police subjects view as ideal.

Spaulding (1948) compared nurse and police applicants with the Kuder Preference Record. He found the police group higher on "social service" interests and lower on "computation" and "clerical" interests. This high interest in "social service" would be found within the concept of the ideal policeman that most citizens and policeman find as desirable.

Kole (1962) and Matarazzo, et al. (1964) reported from the same data comparing medical students and civil service applicants (policemen and firemen candidates). On three measures of emotional adjustment (the Taylor Manifest

Anxiety Scale, the Cornell Medical Index and the Saslow Psychosomatic Inventory) the policemen and firemen applicants showed a "lacking in indices of psychological dysfunction". These authors did not find support for this finding in the Minnesota Multiphasic Personality Inventory (MMPI) results of the civil service group. They report that police applicants had high scores on scales K (defensive responses), Pd (psychopathic deviate), and Hy (hypomania). This pattern of responses was described as "blustery, sociable, exhibitionistic, active, manipulating others to gain their own ends, opportunistic, unable to delay gratification, impulsive, and showing some tendency toward over-indulgence in sex and drinking" (Matarazzo, et al., 1964, p. 132).

These conclusions appear to be in contrast with those of Nowicki (1966). He used the MMPI to compare 27 policemen with a matched group of industrial workers. He found the policemen higher on Cn (which reflects psychological control), p< .10; and higher on Si ("social introversion"), p< .02. He reported that policemen had better control over "overt impulses" and appeared to be "less out-going" than the industrial workers. Both groups (policemen and industrial workers) were more outgoing, however, than test norms for the general population. This fact indicates that these results may be similar to those of Kole (1962) and Matarazzo, et al. (1964), suggesting that the average policeman may be more outgoing ("sociable", "exhibitionistic") than the population norm.

Gottesman (1969) compared MMPI scores of successful police applicants with male veterans appearing for counseling and found that the police group exhibited fewer profile deviations than the veteran group. This result is not unexpected. It would be presumed that policemen would be more stable than those who were seeking counseling.

From these comparison descriptive studies the predominant qualities that emerge show that policemen are very outgoing, people-directed individuals ("exhibitionistic", "sociable") with some immature qualities ("over-indulgent", "manipulative"). Policemen have few debilitating personality qualities (have stability).

Criterion studies: The relating of personality test scores to job performance measures. Criterion-related personality studies attempt to clarify the relationship between personality traits and job success.

> Criterion-related validity indicates the effectiveness of a test in predicting an individual's behavior in specified situations. For this purpose performance on the test is checked against a criterion, i.e., a direct and independent measure of what the test is designed to predict (Anastasi, 1968, p. 105).

The "direct and independent measure" that is of concern in this present research is the job performance score of the policeman.

The MMPI is a personality test that has often been used in police studies which correlated test results with a criterion. Fraser (1949) in a concurrent validity study which involved volunteer R.C.M.P. who were divided into three performance groups by Officers Commanding and personnel officers found that the most efficient group scored significantly higher on the MMPI scales of hysteria and hypocondriasis. From these results and from the entire profile Fraser concluded the following.

> Specifically, the members of the least efficient group tend to be characterized by a less socially oriented (or more inward turning) mode of life adjustment (with more asocial behavior manifestations indicated) than do the members of the most efficient and mediocre groups; of the latter two, it is indicated that the mediocre members are, as a group, slightly less socially oriented than the most efficient members. (p. 7)

These results, based upon concurrent validation, suggest that social skills and some self-centeredness are important for efficiency in police work.

Constrasting results were found by Marsh (1962). He included the MMPI within a battery of tests given to 100 men who had just been appointed deputy sheriffs in the Los Angeles area. (The men were told that these tests were for counseling only.) Excellent criterion measures were obtained. Each supervisor involved was given a set of cards containing the names of subjects he supervised. He was asked to sort these cards into five equal categories from "poorest" or "least qualified" to "best". Over the ten year period different supervisors ranked the subjects. ("The number of ratings per subject ranged as high as seven" (p. 42).) The correlation coefficient for two raters per subject was r=+.51. On the MMPI the hypomanic scale was found to be a good predictor of job performance over the ten year period. "Subjects with T scores below 55 were much more likely to succeed" (p.43). A score below 55 on the hypochondriasis scale was also a good predictor. (It should be noted that Fraser's results (1949) suggested that policemen scoring higher on this scale tended to be more efficient.) Hypomania has been described by Dahlstrom and Welsh (1962) as an affective disorder characterized by "over-activity, emotional excitement, and flight of ideas" (p. 74). Hypochondriasis is a test scale that reflects an abnormal concern with bodily functions. Low scores on these two scales would indicate stability, calmness, and lack of self-centeredness. It appears form Marsh's study that these qualities, which are reflected by low scores on hypomania and hypochondriasis, are important prerequisites for above average future job performance of policemen. The advantage of Marsh's study over Fraser's is that Marsh's study has predictive validity over ten years. Therefore, Marsh's results should be given a greater weight. However, the weakness in Marsh's study relates to the testing session. The subjects, as noted previously, were informed that the tests were for "counseling purposes only". This means that these subjects might have been more open and frank in this situation then in a recruitment assessment situation. Similarly Fraser used volunteers who were assured that the results would not be recorded on the their files. It would be erroneous to generalize these two studies to an assessment situation where applicants would be answering questions (in a recruit situation) so as to appear more acceptable for the position of policemen.

As mentioned the advantage of Marsh's study is that the MMPI test variables were correlated with later on-the-job performance. When this approach is not used, and the tests have not been validated for the situation within which they will be used, interpretation of the test results is more difficult. For example, Rhead, Abrams, Trosman, and Margolis (1968) were involved with the Chicago Police Department in selecting police candidates. They reported the following after obtaining the MMPI profiles of over a thousand police applicants.

> The group profile of the candidates exceeded the average on the Pd (psychopathic deviate) and Ma (hypomania) scales, scales correlated with a willingness to take chances and with a propensity for acting out. (p. 1578)

The results of Rhead and his colleagues (1968) point out that the average Chicago police candidate is scoring highly on a MMPI scale (hypomania) that may be predictive of poor job performance (as indicated by Marsh, 1962). Obviously this occurrence makes clear that descriptive studies of police characteristics are of less value to selection officers (because they are not related to actual future job performance appraisals) than the predictive studies.

Blum (1964) conducted a major study involving a seven year follow-up of a group of 87 systematically tested recruits. The test results were not used in the selection decision and Blum attempted to predict the subjects' job performance on a three point scale; one-satisfactory, two-questionable, three-disciplinary or efficiency problem. Blum had no knowledge of the applicants' background. After the policemen were on the job for seven years Blum found that the Schizophrenic Scale and the F Scale (unusual responses) correlated r=+.20 or higher with negative performance measures ("exceptional serious misconduct", r = +.47, r = +.42). These correlations over a seven year period indicated that some negative job behaviors were predictable with the use of certain personality variables.

Hooke and Krauss (1971) compared the MMPI test scores of successful candidates for promotion to sergeants with a matched group of patrolmen of the Kansas City, Missouri, police department (37 policemen were tested). The authors reported the following descriptive information.

> The MMPI profiles of both the group of successful candidates and the comparison patrolmen peaked on scales 4 (psychopathic deviate) and 9 (hypomania) .... Overall, so far as the MMPI is adequate to reflect psychopathology, both groups of police officers are a normal lot, exhibiting high energy and little neurotic inhibition. (p. 105)

Significant differences between the two groups occurred on three scales. The successful candidates were higher on scales K (defensiveness), Pa (paranoia) and lower on scale O (social introversion). The authors concluded the following.

> These differences suggest that as a group the successful sergeant candidates tend to depend more upon themselves, appear more self-confident, are more sensitive in interpersonal relationships, and are more outgoing and genial than the patrolmen. (p. 105)

The results of these MMPI studies show only one repeated finding. The average profile of police candidates peak on the scales of "psychopathic deviate" and "hypomania" (Matarazzo, et al., 1964; Rhead, et al., 1968; Hooke and Krauss, 1971). The predictive relations reported have not been replicated. The best studies have indicated that low scores on "hypomania" (Marsh, 1962), "hypocondriasis" (Marsh, 1962), "schizophrenia" (Blum, 1964), and the "F Scale" (Blum, 1964) would be predictive of good police performance after seven to eight years on the job. Hooke and Krauss (1971) have reported that high scores on Scale K (defensiveness) and the scale indicating "paranoia" were concurrently predictive of successful sergeant candidates. These contrasting results among the studies reviewed make it difficult to determine which personality variables are important in predicting future police performance. The differences that occurred may be a reflection of different police departments, different concepts of ideal policemen, or different ways of measuring job criteria. From these studies, with the lack of replication, few firm conclusions can be drawn concerning the predictive value of the MMPI.

A limited number of studies report predictive results with other personality tests. Hogan (1971) gave the California Psychological Inventory (CPI) (Gough, 1969) to three groups of police cadets and one group of experienced (one year) policemen who were taking a refresher class. "Staff members at the Police Academy were asked to rate each cadet for 'overall suitability for police work'" (p. 681). The disadvantage of this type of criterion is twofold. First, it is a concurrent type of validation and may not be predictive of future performance. Second, the performance rating is based on the instructors' viewpoints after observing the subjects' classroom behavior. Classroom performance may have little relationship to on-the-job performance. As one might expect, the "Intellectual Efficiency" scale was the best predictor (correlated most highly with classroom performance) for three of the four groups. This scale is described in the CPI test manual as indicating individuals who are "...planful, thorough, and resourceful; as being alert and well informed; and as placing a high value on cognitive and intellectual matters"

(Gough, 1969, p. 11). This test scale is in general one of the best predictors on this test of classroom behavior. Other variables that predicted the criterion fairly well were "Achievement via Independence" and "Well-Being". Hogan (1971) suggested the following conclusion. "The dominant theme is the importance for effective police work of practical functional intelligence in combination with sociability and self-assurance" (p. 684). This conclusion is a generalization which goes beyond the limitations of the criterion (classroom performance) used in the study. For this reason the conclusion is questionable. This research should be replicated with a more representative on-the-job criterion.

Thweatt (1972) used the 16PF (Sixteen Personality Factors) (Cattell, Eber, and Tatsuoka, 1970) to compare 50 experienced police officers with a group of men that had left the police force. The police group scored significantly higher than dropouts and men in general on Factor G, "super ego strength". The dropout group were higher than the police group on Factor I, "tender-heartedness". These results are confounded by the fact that three groups were tested (experienced norm group, n=50; recruit dropout group, n=18; recruit remain group, n=87) while statistics were reported on only two groups (norm and dropout groups). The significant differences that occurred between the norm group and the dropout group may not have occurred between the dropout group and the remain group.

These omitted facts on the part of the author lessen the value of these findings.

Other research with apparently good results has been conducted with obscure tests. Humm and Humm (1950) used the Humm-Wadsworth Temperment Scale in predicting which policemen in the Los Angeles Police Department would fail to remain on the job. They indicated a correct prediction of 91% for the the fail group. Blum (1964) discusses this article and has noted the following.

> It appears that among a total of 669 men tested, 79 resigned, 233 were fired, and 357 remained with the department. The ratio of men fired to men resigned is most unusual and suggests a very special situation.... (p. 107)

Fifteen years later Collins (cited in Blum, 1964) compared Humm-Wadsworth scores in the same police department with a criterion measure and found no significant results.

Morman, Hankey, Heywood, and Liddle (1966a) tested 62 Kansas State Traffic officers with the TAV Selection System. This relatively unknown test is based on Horney's theory of personality (T = move toward people, A = move away from people, V = versus people). The criteria used were five supervisors' ratings of each traffic officer's performance. A multiple R of +.51 resulted when test variables were correlated with the criterion. In another study (1966b) of police recruits where performance during 16 weeks in the police academy was used as the criterion a multiple correlation of R = +.46 was reported. In a later study (1967) a multiple correlation of +.90 was found between nine test variables, three non-test variables (age, education, experience) and the supervisory rating criterion. It should be noted that with this research the amount of time between the testing session and the criterion assessment was not clearly stated. This means that the results have to be considered as concurrently validated. Their predictive validity was not indicated. These results suggest that certain personality variables are correlated (at least concurrently) with job performance. The statistical approach of multiple correlation is a valuable one for it allows the consideration of a large number of variables in discriminating between good and poor job performers.

It should be pointed out that not just any personality variables will differentiate between successful and unsuccessful policemen in a multiple correlational study. Hankey (1968) tested 801 policemen appointed to the Los Angeles Police Department from 1955 to 1959. He reported his procedure as follows.

> The predictors were ten trait and two falsification scores from the Guilford-Zimmerman Temperament Survey, scores form the California Test of Mental Maturity, and the Wonderlic Personnel Test, the rating on the civil service entrance examination; and age, education, military rank, occupation and law enforcement experience at time of appointment. The criteria were three measures of recruit academy performance, a work performance rating, three measures relating to commendatory behavior, the rating on the oral portion of the sergeants' examination, and four

items relating to job tenure. (p. 952)

From this apparently well conducted study Hankey reported that: "No evidence was found to support the hypothesis that successful policemen show a different syndrome of personality traits or other variables than nonsuccessful policemen" (p. 952). The instructions given to these men before being tested are not stated. It is presumed that these subjects had already been accepted into the police department and therefore their test results may have had no effect on tenure considerations. Under conditions such as these, the subjects may not have considered the testing in a serious manner.

From this review of personality variables and police performance very few substantial conclusions can be drawn. Only one study has been replicated, that of Morman, et al. (1966a, 1966b, 1967). This study was not predictive but concurrent in nature and relatively unknown test variables were used (TAV Selection System). The other predictive studies which used the MMPI personality scales offered good results that were not replicated in similar studies (Fraser, 1949; Marsh, 1962; Blum, 1964; Hooke and Krause, 1971). One study shows no relationship between personality and job performance (Hankey, 1968). Obviously a more through and long term approach to this problem of the relationship between personality characteristics and the job performance of policemen is required in research.

Policemen and Stress

In selecting policemen personnel officers look for men who can withstand the stress that is part of the job. It is reported that policemen have a significanatly higher rate of ulcer and digestive tract problems (United States Department of Health, cited in Earle, 1973; Los Angeles County Sheriff's Department, 1971). Lewis (1973) and Neiderhoffer (1967) both have pointed out that policemen in New York city have a higher suicide rate than the general public in that city. Nelson and Smith (cited in Heiman, 1975) have made similar remarks regarding the American state of Wyoming. They report that the 1960-1968 suicide rate for policemen was 203/100,000; "almost twice the rate of physicians, the next highest group" (p. 295). Heiman (1975) in another article on suicide compared the high police suicide rate in New York with the police suicide rate in London. Significant differences were found. He also pointed out that the London police suicide rate was virtually no different than its white male urban population.

The above results suggest that policemen in North America should be able to cope with events that may lead to suicidal thoughts. Specifically they must be able to develop police behaviors required for the role of law enforcer, public servant, etc. When learning these roles, or when not capable of fulfilling them, high degrees of stress may result. It follows that candidates for positions of policemen should be screened not only for their ability to handle the stress of the position but also for indications of past psychosomatic reactions to stress.

Situation Tests and Police Selection

Many studies concerning the assessing of candidates whose later performance will require intelligent action under stress have used situation tests to select these candidates. Chenoweth (1961) encouraged this development in police recruitment in an article titled "Situation tests: A new attempt at assessing police candidates". This approach is relevant when considering potential police candidates because of the importance for policemen of maintaining emotional control under stress. Rhead, Abrams, Trosman, and Margolis (1968) have emphasized the importance to the assessment officer of knowledge regarding each potential policeman's means of regulating aggression.

> A matter of particular concern for us was the applicant's ability to use aggressive energies in effective and adaptive channels. The police officer finds himself in occupational situations necessitating reality adapted action patterns. Thus the capacity to regulate and control aggression in terms of appropriate goals was given a high weight in the selection process.
>
> ...We tended to be concerned about those applicants who manifested primitive uncontrolled aggression leading to disorganized behavior in socially maladaptive patterns. (p. 1579)

The situation test is one approach that allows the police candidate to demonstrate his emotional control in a stressful situation.

Moreno (1949) was the first to use the term "situation test" and did so in 1932. At that time he described the situation test as "...a method of testing which is patterned after a life situation" (see Moreno, 1949, p. 349). During a situation test the candidate would be placed in a job-related situation so that his reaction to that situation could be observed. Chenoweth (1961) has defined situation testing as "...a technique through which a candidate is exposed to a carefully constructed situation, his reaction to the situation produces behavior that may predict his reaction to comparable situations in the future" (p. 234).

One advantage of the situation test over other tests is that it is difficult to fake. Flannagan (1954) has reported that when a person is questioned (on a paper and pencil test or in an interview) about his performance in a particular situation he will usually modify his answer in the direction most advantageous for him. In a situation test the subject's response is in terms of more than just verbal or written behavior. He has to call upon behavior patterns that he has used in the past that involve physical movement, expression of emotion, the coping with stress, leadership responses, and problem solving. He is allowed to try to do his best in dealing with the task he is presented. His reaction within the situation test will demonstrate the appropriatness of the behavior patterns that he has learned and thinks are applicable in that situation. This is one of the main advantages of the situation test. It allows the candidate to honestly demonstrate his potential to do the job by trying to represent himself to the best of his ability.

Another important advantage of the situation test is that it allows for the observing of behavior that occurs infrequently.

> The greatest advantage of the situational test observation is that it makes possible the observation of characteristics which appear only infrequently in normal activities ...characteristics such as bravery, reaction to frustration, and dishonesty. A single situation test may reveal more about such a trait than weeks of field observation. (Cronbach, 1949, p. 414)

This advantage is especially important in the assessment of police candidates. It is information about these very qualities (bravery, frustration tolerance, honesty) in police candidates that personnel officers have difficulty in gathering accurate information.

Bersoff and Greiger (1971) point out disadvantages of paper and pencil tests and in so doing cite furthur advantages for situation testing. They state that tests administered in a class room or clinical setting have limited usefulness in predicting overt behavior because the testing environment is designed to draw out the best performance of a subject under ideal conditions (quiet surroundings, low stress level, continual positive reinforcement from the examiner). For example the situation surrounding an intelligence testing setting is not representative of the situation in which the subject will be called upon to use his intelligence. The authors call for "psychosituational assessment" where behavior is measured and interpreted in the natural or contrived natural situation. This should then be more predictive of future behavior in similiar situations.

The content of the situation test itself is another problem that has to be considered. Cronbach (1956) has stated that assessment is easier when "...a test is a sample of a criterion" (p. 185). McClelland (1973) has agreed with this point and states that the best testing is "criterion sampling". The content of the situation test should be closely related to important aspects of job performance. "Fundamentally, it is typical performance on the job which the situation test should measure" (Weislogel, 1954, p. 493). The situation test should allow the police candidate to demonstrate three things related to the job in question: (1) possession of the necessary skills, (2) recognition of the need to apply these skills, (3) motivation or willingness to apply these skills (see Weislogel, 1954). Once the police candidate's behavior has been demonstrated in the job related situation the assessment of this behavior in terms of predicting future job performance should be easier because of the high similarity between test content and job content.

The measuring of situation test behavior has to be considered in terms of standardization and reliability. The standardization of a situation test is often cited as one of its weaknesses. Weislogel (1954) has stated the essential elements of standardization for situation tests.

> Problems (the situations) should be structured so that each subject faces the same critical situation. This is an obvious, but sometimes difficult requirement to meet in devising situation problems for groups of subjects. It is essential that the same need for action be present for all participants. (p. 496)

With good standardization each candidate will be presented with the same situation. This approach makes it much easier to compare candiates.

The scoring of situation test performance is another difficulty. Raters are often used to assess the candidates behavior. Situation tests tend to be characterized by the fact that "... agreement between scorers or observers is less than perfect" (Fiske, 1954, p. 465). This result occurs for two main reasons. First, the content of things that the subject manipulates within the test are real, for example, situation tests often require subjects to interact with other subjects or with actors. This lessens the control that the experimenter has over the stimuli with which the subject interacts. Second, situation tests are used not to measure pure traits, but are used in assessing complex skills such as leadership or tolerance to stress. This means more variance is involved in situation testing then in other forms of testing. For these reasons reliability between raters may be lowered and rater agreement lessened. With this problem of reliability situation tests should be designed to discriminate among individuals on variables that will be widely differentiating.

Many assessment batteries have included situation tests. The use of these situation tests will be reviewed so that the applicability for police candidate assessment can be considered. Examples of issues previously discussed will be pointed out to clarify the advantages and disadvantage of situation tests. These issues will include the following: the assessment of the emctional control of the subject, the advantage of allowing the candidates to try their hardest, the potential within the test to observe infrequent behavior , the use of the test as a 'criterion sample' (a measure of job behavior), the problems of standardization and reliability with this type of testing.

Situation testing came to prominence during World War II but its beginnings can be traced tc earlier periods in time. It is reported in the Bible, Judges 7:4-7, that Gideon used a type of situation test to pick out those of his people who would stay and go into battle with him. He asked all of his men to go down to the river and drink. He chose the men who drank from the river using their hand to hold the water rather than bending from their knees and putting their mouth directly in the water. Although the rationale for this test is not clear, Gideon did choose those who drank in the most defensible position. In the confrontation that later ensued Gideon and his men were successful, partially validating the situation test.

Galton (1884) foreshadowed the use of situation tests. He believed that character should be measured by noting definite acts in response to particular situations. These situations he felt could be contrived: "Emergencies need not be waited for, they can be extemporized; traps as it were can be laid" (p. 182). The situation test that he developed was based on the assumption that friends tend to slope together when sitting side by side. He devised a pressure gauge to be placed on the legs of a chair to investigate this hypothesis. Although no results were reported, this serves as an example of a good situation test where behavior is being measured (how the person is inclined in the chair) rather than measuring the subjects verbal response (how he reports he feels about the other person).

Fernald (1912) designed a situation test in an attempt to differentiate between normals and delinquents. He required his subjects to stand on tiptoes with heels 1/8 inch or more from the floor. This test significantly discriminated between the groups (the normal group stood on tiptoes longer than the delinquents). Unfortunately the testing situation was too long to be pratical. Some subjects were spending over 50 minutes on tiptoes. (This test could have been shortened in length by having the subjects hold a weight during the testing.)

Aggressiveness was measured by Moore and Gilliland (1921) with rather a unique situation test. They believed that a steady gaze was characteristic of the man of power. Their subjects were required to perform a series of mental additions while constantly returning the fixed gaze of the instructor. It was found that men rated high in aggressiveness made fewer eye movements to avoid the stare of the instructor. Although in this study the situation test is a primitive measuring tool it demonstrates that actual interpersonal behavior is being assessed. However, this test is not judged altogether representative of actual situations where aggression would be expressed. It is possible that the more distant a testing situation is from the actual situation the greater liklihood that factors that function better under little stress will control the outcome. For this reason it is necessary to keep the testing situation as representative of the actual situation as possible; a 'criterion sample'.

Hartshorne and May (1928) were requested by religious organizations to investigate the morality of children. What resulted were ingenious situation tests that measured "honesty", "trustworthiness", "helpfulness", "inhibition", "persistence", "moral knowledge" and "attitude" (also see May and Hartshorne, 1925; May and Hartshorne, 1926). One example of the situations they developed for young children was the following. In Part 1 of a test they would ask questions such as: "Do you know who discovered America? Yes/No." After this section of the test was collected Part 2 was distributed and included questions similiar to: "America was discover by: Drake, Columbus, Balboa, Cook?". This measure of honesty also discriminated between those who adjusted well and those who adjusted poorly. In this situation test, behavior demonstrating honesty or dishonesty is observed rather than examining the subject's responses to questions directed toward discovering his opinion on his own honesty. Situation tests should have the advantage of being more representative of actual behavior than paper and pencil tests. In this study, infrequent behavior was observed (dishonesty) with the use of 'criterion sampling'.

Of more relevance to the purpose of this research project with the police department are the situation tests carried out within a military setting. The German psychologists devised situation tests to select potential officers during World War II. Ansbacher (1941), Ansbacher and Nichols (1941), and Ansbacher (1949) have reported that the following qualities were seen as desirable in German army officers: "imagination and rapid learning ability, capacity for swift adjustment, initiative and will power in thought and action, and emotional stability and security of conduct" (Ansbacher and Nichols, 1941, p. 44). The German psychologists tested four main areas to investigate these criteria.

(1) For the Intelligence analysis the candidates were required to write a composition about a picture. They were also asked how they would organize a certain type of meeting. Furthur questions were asked about how certain military commands would be carried out. This section of the German testing could not be described as situation testing. These tests mainly required a verbal report on the action the candidate thought he would undertake in a situation rather then having the candidate act through a contrived situation. His thinking and organizational potential would be demonstrated with this type of testing.

(2) The Action Analysis section of the test battery contained some situation tests. In the "command series" candidates were required to walk along a horizontal pole, with pack on back, securing ropes to high hanging hooks. Scaling a wall was also involved. As part of the test the candidate was continually criticized. A major factor in the assessment of this test was how easily the candidate became discouraged. His attitudes when approaching tasks while under stress were assessed carefully.

(3) The Expression Analysis portion of the testing also

contained situation tests. The candidates were required to instruct a group on an unfamiliar task (e.g., how to make a coat hanger). "Warmth" and "sociability" were two variables which were considered important. The candidates' speech qualities were also observed during this portion of the testing for "loudness, pitch, melody, timbre, articulation, tempo and pauses".

(4) The Life History portion of the testing involved an interview concerning the background of the candidate and the answers to certain questions (e.g., "What sort of people does he prefer?", etc.).

The main criticism of the German approach has been an often cited disadvantage of situation testing. Because situation tests are lifelike it is more difficult to control incoming stimuli. This lifelike quality of the situation tests, allows the range of responses of the subject to be very broad. These characteristics (lack of control of all stimuli and the possibility of a wide range of responses) necessitate careful observance of guide lines in terms of standardization, uniformity of procedures, and consistent scoring approaches. These areas of testing were the weakest in the German approach.

Fitts (1946) reported on this weakness. He visited German universities and interviewed German psychologists following World War II. He made the following comments concerning their assessment procedures.

> A marked characteristic of the German testing procedure was the emphasis on subjective evaluation of test data. German psychologists stated repeatedly that observations of the candidate's behavior during a test were more important than the actual score which he earned. ...Individual examiners were permitted and often encouraged to vary testing procedures and to emphasize their favorite tests. ...The concepts of reliability and uniformity of method appear to have been lacking. (p. 153)

Jenkins (1947) came to a similiar conclusion. "The German program is a beautiful example of the uselessness of elaborate testing techniques hand in glove with complete disregard of their necessary concomitants---standardization, objectification, and validation" (p. 69). These deficiencies in the German approach have to be guarded against in the designing of situation tests. Often the assessor assumes that because his test is lifelike it must be valid for predicting future behavior. Because of the belief in the validity of the test, standardization and reliability procedures are often forgotten and the validity of the test can never be established because of the lack of precision in administering and scoring the tests.

In considering situation tests, most criticisms are directed at the disadvantages already cited. (Low reliability and validity measures occurring as a result of poor standardization and scoring techniques.) It is important to remember Cronbach's (1970) definition of a test: "A systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category system" (p. 26). This is possible within the framework of the situation test.

During World War II the Americans and British also used situation tests. Murray and his colleagues (1948) assessed armed forces personnel who were candidates for especially dangerous missions behind enemy lines (also see Murray and Stein, 1943; Murray and MacKinnon, 1946). Their group-situation tests included the "Wall Test" (a group of men were required to get over two parallel walls with their mock bazooka) and the "Discussion Test" (where six men discussed a question for half an hour). These tests were designed to draw out leadership abilities which could emerge in these unstructured group situations.

The individual situation tests were comprised of the "Construction Test" and the "Stress Interview". For the construction test the candidate was required to direct assigned helpers (who were very uncooperative) to build an object. The stress interview involved having the subject assume that he had just been caught going through a secret enemy file and as a result was apprehended and grilled concerning his motives

These tests were directed towards discovering qualities of leadership and reaction under stress. Although no conclusive results were reported a large number of men were directed away from dangerous missions based on their performance during this testing. The opportunity to observe characteristics such as those mentioned is the advantage of the situation test. It allows for the assessment of characteristcis which seldom appear in normal activities.

Policemen may also find themselves in highly stressful situations where they must act with leadership and intelligence. In test situations that measure leadership and intelligence independantly of stress, the test results may not be representative of the candidates actual behavior in situations involving stress.

Following World War II situtation tests were given in assessment batteries with limited success. Kelly (1954) gave situation tests designed to evaluate beginning graduate students in clinical psychology. As stated previously the situation test has been best used in assessing characteristics such as bravery and interpersonal reaction under stress, characteristics infrequently seen in normal behavior. These are important qualities in any individual, but may not be considered essential qualities in determining the success or failure of graduate students. The tests given the subjects included the following: a leaderless group discussion where a topic is suggested and without leadership the group discusses the topic; role playing by two candidates interacting with each other; nonverbal expression of emotion (subjects were required to react to an emotionally laden word in a nonverbal manner); a group task involving heavy physical labor; a staff-candidate party.

The results of these tests were generally poor, not predicting as well as a battery of tests. One reason the tests failed was because of the poor standardization. In none of the situations were the candidates assessed alone. For example the improvisation test involved two candidates role playing with each other. Therefore each candidate was in a unique situation with another candidate. If one candidate took the required action to solve the problem presented, the "need" no longer existed for the other candidate to take that action. Therefore the second candidate was in a different situation than that of the first candidate.

Another problem with the research by Kelly was that aspects of typical job performance were not being considered. How representative of typical performance of a graduate student is the nonverbal expression of emotion? The deficiencies in this study concerning the assessment of graduate students must be avoided in situation testing. Standardization and "criterion sampling" (representativeness of job behavior) are essential.

Situation tests have been used occasionally for police recruit selection. Mills, McDevitt, and Tonkin (1966) included three situation tests in their test battery. The "Foot Patrol Observation Test" was comprised of a six block walk through a part of Cincinatti according to written instructions followed by a questionnaire which measured their observational ability. For the "Clues Test" an office setting was laid off in the police gymnasium with clues planted around the desk so that certain assumptions could be postulated about the living habits and possible whereabouts of a city employee. Candidates were told to investigate the mysterious disappearance of this person. Following ten minutes of investigation they were given a questionnaire to complete. The "Bull Session" was the final situation test. Eight to ten candidates were assembled following a dinner break and offered coffee and cigarettes. Two group leaders and two observers assessed the candidates while they were encouraged to discuss with their peers and group leaders such relevant topics as: "use of force, the handling of fear, alcoholism, use of narcotics, mental illness, prostitution, homosexuality, administration of justice through the courts, minority groups, and the use of authority". During these discussions, the candidates' abilities for peer group interaction and their attitudes towards different aspects of police work were assessed.

Results of two of these situational tests were reported by the authors (the Bull Session results were not included). The criterion was established by dividing the candidates into a success and failure group. The 'Success Group' was comprised of the forty-two candidates who completed their training while the 'Failure Group' was made up of twenty candidates who did not complete the training. The results are shown in Table 7. Results indicated that these tests

Table 7

Test Performance of Success and Fail Candidates
(from Mills, McDevitt, and Tonkin, 1966)

Test	"Success" Group Mean (n = 42)	"Failure Group" Mean (n = 20)	Mean Difference	t*
Foot patrol	14.7	14.0	.7	.170
Clues	31.2	26.2	5.0	.597
AGCT	125.5	122.5	3.0	.090

*None of the t tests showed significant differences.

did not contain variables that significantly discriminated between the two groups. But when the test scores of the original 'success' group were rank-ordered according to academic standing in training school the correlations as presented in Table 8 occurred. The intelligence score of the AGCT predicted police academy standing quite well. The clues test significantly predicted academy standing, but it is a low level of prediction. (For a later group of successful candidates this correlation was not significant.) What is of interest here is that the clues test was not highly correlated with the intelligence measure (r=+.105) and yet was correlated to a certain degree with academy standing (r=+.375). The authors' plans for refining the situational tests involved having a measure independent of paper and pencil intelligence tests which could be developed to predict successful job performance. If that could be done, two measuring devices, situational tests and intelligence tests could be used to assess two different job requirements, namely, basic intelligence and job competence. (It should not be overlooked however, that the criterion in this research is weak. Training school marks may not be representative of job performance.)

Although the study did not report succinct statistical results on the "Bull Session" portion of the testing the authors did make generally favorable comments. They stated that unsuspected character traits and attitudes appeared in the group discussions. It was also noted that some men who "had previously performed adequately on paper and pencil tests "...became disorganized, rambling and circumstantial" during the "Bull Session". It was stated that "...the Bull Session was the most valuable technique used in recruit selection" (p. 104).... This conclusion can be challenged. The results as reported for this portion of their test battery are subject to all the criticisms of the German Military testing that were previously cited. Basically these criticisms include the lack of standardization and validation for the testing session. Suspected good results have to be substantiated by ascertaining the accuracy of predicting performance based on the testing data. High correlations between test results and performance can only be accepted as accurate if testing and scoring procedures are standardized. By excluding validation and standardization from this situation test the "Bull Session" loses most of its value.

Table 8

Rank Order Correlations of Successful Candidates
(from Mills, McDevitt, and Tonkin, 1966)

	Foot Patrol	Clues Test	AGCT
Foot Patrol	----		
Clues Test	.099	----	
AGCT	.211	.105	----
Police Academy Standing	.137	.375*	.595**

*p<.05
**p<.01

In summary, a good situational test might be a very effective means of assessing police candidates. Certain advantages have been pointed out. Candidates may try their hardest and not worry about presenting themselves dishonestly ("faking good" is not a problem). Specific job skills are tested in job-related situations to assess candidates rather than trying to predict ability on job performance from paper and pencil test scores that were attained in a non-stress situation. Characteristics required for police work are difficult to ascertain in a classroom setting. The qualities of good commonsense under stress, bravery, and appropriate expression of emotion are all important in police work and are probably best obtained with situation tests. By its very nature the situation tests is more difficut to standardize and make reliable. If situation tests are to included in a test battery special caution must be directed towards ensuring that good standardization is employed and reliability achieved.

Research Hypotheses

Intelligence

Tests to be used: Wonderlic Personnel Test, Form B, Wonderlic, 1942; Wechsler-Bellevue Form I, Vocabulary, Wechsler, 1946.

1(a) Policemen will fall within the average to above average range of intellectual functioning on the Wonderlic Personnel Test and the Wechsler-Bellevue vocabulary test.

1(b) Little relationship between intelligence scores (of this above-average intelligence group) and performance measures has been demonstrated. It appears that when a group of police candidates are at an above average intellectual level, it is variables other than intelligence, that determine job performance abilities. For this reason no hypothesis is made concerning the discriminatory power of this measure. This test will be included primarily to establish Canadian norms for the intellectual levels of recruit applicants.

Interests

Test to be used: Strong Vocational Interest Blank, Strong and Campbell, 1966.

2(a) The three groups of policemen (high level, average level, and low level performers) will fall within the 'A' category on the "policeman" scale. This scale will not discriminate among performance levels.

2(b) High performing policemen will show a higher interest on mechanical and clerical measures. The scales of "mechanical" and "technical supervision" will be observed to indicate if these differences are present.

2(c) High performing policemen will tend to score lower on the professional scales than the other two police groups. The scales of "dentist", "osteopath", "veterinarian", "biologist", "architect", "physician", "psychologist", "psychiatrist", and "author/journalist" will be observed carefully to note this discrimination. The scale "diversity of interest" which measures "breadth of interest" is also predicted to show that high level policemen do not have as wide a range of interests as do the average and low level performance groups.

2(d) It has been observed in the literature (Berman, Darley, and Patterson, 1934, p. 228; Matarazzo, Allen, Saslow and Wiens, 1964, p. 129) that high performing policemen have a strong interest in social contacts. It is hypothesized that the better-performing policemen will score higher on social contact scales (e.g., "personnel manager") than poorer-performing policemen.

2(e) It is hypothesized that policemen will score highly in comparison with population norms on the "masculinity-femininity" scale indicating "typically masculine" interests. As mentioned earlier, Johansson and Flint (1973) have noted this occurrence with police groups and have interpreted high means scores on this scale as "indicative of strong authoritarian-structured types of interests" (p. 42). No hypothesis concerning the discriminatory power of this variable in regard to different performance levels of policemen is made because none has been previously demonstrated.

Personality

Tests to be used: California Psychological Inventory (Gough, 1969); Thematic Apperception Test (Murray, 1944); Draw-a-Person (Machover, 1949); Cornell Index Form N2 (Weider, Wolff, Brodman, Mittleman, Wechsler, 1948); Hostility and Direction of Hostility Questionnaire (Caine, Foulds, and Hope, 1967); Incomplete Sentence Blank (Rotter and Rafferty, 1950).

3(a) For the projecteve tests (Thematic Apperception Test, Draw-a-Person, Incomplete Sentence Blank) the possibility of "faking good" is diminished. It is hypothesized that the high-performing policemen will respond to the test items in a more mature manner than will the low-performing policemen. The mature perspective shown in the subjects responses will be rated according to categories defined in the scoring procedures for each test (see Appendices 2,3,4,5).

3(b) It is hypothesized that on the three more standardized tests (California Psychological Inventory, Cornell Index Form N2, Hostility and Direction of Hostility Questionnaire) the recruits' means will fall within the normal range. Because of the possibility of "faking good" on these tests, and also because the literature has not demonstrated the discriminating ability of these tests for this type of prediction problem (discriminating high level and low level police performers), no hypothesis in this area will be formulated. These tests will be included primarily to establish Canadian norms for the personality variables that they contain.

Situational Tests

Tests to be used: Situation Tests and Three-Minute-Tests.

4(a) It is hypothesized that the high performers will respond to the test situations in a manner that demonstrates good "commonsense", "courtesy", and ability to handle stress in an interpesonal situation. For these test variables (as measured by "approach to situation", "clarity of verbal request or response", "consideration given to other's comments", "effectiveness of solution to problem", "confidence shown in handling situation", "interpersonal skills", and "level of development") the high-performers will be given ratings indicating a more mature level of development than the low-performing policemen.

Overall Prediction

All test variables considered.

(5) It is predicted that when all the test variables are considered, and the relevant variables combined in a multiple discriminant analysis, a significant discrimination among three levels of police performance will result. Once this has occurred a discriminate function (or functions) based on the analysis will be formulated which will allow for future predictions of level of job performance based on the psychological test scores.

CHAPTER II

METHOD

Subjects

The subjects in this research were two police recruit classes of the City of Edmonton Police Department. The sixty-one (61) police recruits were tested within the first week of their training clasess (Class 37, August 7-10, 1973; Class 38, October 3-8, 1973). Of the original 61 police recruits, 58 were males, 3 were females. The average age of the subjects was 23.26 years. Police applicants at this time had to be 5' 10" in height, be at least 150 pounds in weight (but not be over-weight), and have a grade 12 education or equivalent. Upon evidence of these basic requirements candidates were interviewed, their references were verified, and they were given a written examination (which included tests of general knowledge, arithmetic skill, grammar, and composition). Following this they were required to pass a medical examination (which included a vision test). Those accepted up to this point were considered by a selection board.

Table 9 shows how the final recruit class members were chosen. As shown in this table the subjects in this research project are a highly select group. A large number (322) of those initially applying (833) did not return their application form. (They may have realized that they would not meet certain requirements.) The importance of the written exam and the personal interview should be noted. Thirty-seven percent of applicants meeting most basic requirements up to this later stage in the acceptance process were rejected because they failed the written exam. This form of screening will affect the average test scores of intelligence measures used in this research project. Similarly, seventeen percent of remaining applicants were rejected because of the personal interview. This also affects the results of tests used in this research project that sort candidates' scores along dimensions similar to those considered in the interview. It can be concluded that the sixty-one candidates that took part in this study are highly selected and are primarily from Alberta. The effect of these selecting characteristic should be considered when the results of this study are being generalized to other police ~~departments~~ department.

Table 9

Selection of Recruits for Class 37 and Class 38

	Alta.	B.C.	Sask.	Man.	Other	Total
ENQUIRIES RECEIVED	679	31	57	35	31	833
Rejected on Initial Enquiry Because of Failure to Meet Basic Requirements	186	2	5	4	2	199
Application Forms Issued	493	29	52	31	29	634
Application Forms Not Returned	230	18	41	23	20	332
APPLICATION FORMS COMPLETED AND RETURNED	263	11	11	8	9	302
APPLICATIONS REJECTED:						
Lack of Necessary Education	7	-	1	-	-	
Under Minimum Height Standard	12	2	1	-	-	
Under Minimum Weight Standard	4	-	-	-	-	
Over Maximum Weight Standard	6	-	-	1	2	
Not Meeting Vision Requirements	33	2	3	2	-	
Other Medical Reasons	-	-	-	-	-	
Age Requirements	5	-	-	-	1	
Adverse Information on Application Form	1	-	-	-	-	
Applicants Who Withdrew Application	6	1	-	2	1	
Applications Received Too Late to Process	27	2	3	2	4	
Applicants Rejected to This Stage	101	7	8	7	8	131
APPLICANTS STILL ELIGIBLE	162	4	3	1	1	171
APPLICANTS WHO REPORTED FOR TESTS	142	4	3	1	1	151
APPLICANTS WHO FAILED AS INDICATED:						
Minimum Height When Measured	-	-	-	-	-	
Over Maximum Weight when Weighed	-	-	-	-	-	
Written Examination	52	1	-	-	-	
Personal Interviews	23	1	1	-	-	
Medical Reasons	4	-	-	-	-	
Background Enquiries	2	-	-	-	-	
Deferred	1	-	-	-	-	
Applicants Who Failed Tests To This Stage	82	2	1	-	-	85
APPLICANTS STILL ELIGIBLE	60	2	2	1	1	66
APPLICANTS REJECTED BY SELECTION COMMITTEE	8	-	-	-	-	8
SUCCESSFUL APPLICANTS	52	2	2	1	1	58
APPLICANTS ACCEPTED FOR CLASS #37	23	1	1	-	-	25
APPLICANTS ACCEPTED FOR CLASS #38	24	1	1	1	1	28
APPLICANTS ACCEPTED AS CADET CONSTABLES	5	-	-	-	-	5
APPLICANTS DECLINING TO REPORT FOR TRAINING	-	-	-	-	-	
CADET CONSTABLES ENTERING CLASS #37	8	-	-	-	-	8
CADET CONSTABLES ENTERING CLASS #38	1	-	-	-	-	1
*TOTAL CLASS #37	31	1	1	-	-	33
TOTAL CLASS #38	25	1	1	1	1	29

189 letters were sent out to persons who had made enquiries prior to the recruiting drive.

* One was a former member who was re-employed and transferred directly to Patrol Division for employment.

Average Age of Class #37 - 23.36 Years

Average Age of Class #38 - 23.17 Years

Description of Psychological Tests, Administration and Scoring

A number of psychological tests were administered to the subjects. Paper and pencil tests were given in group form during the first two periods of testing. After this, situation tests were administered to each subject individually. The testing took place within the first days of the training program so that the instruction received during the training would have minimal influence on the test performance of the subjects.

The subjects were told at the beginning of the testing that: "These test results are primarily being considered for research purposes but they may be also used in considering the progress of each police recruit". This statement was included so that these results could be generalized to the test battery when used on future occassions for applicants whose acceptance into the police department would be more dependent upon these test results then was actually the case in this research.

Raters Used in this Research

For a number of the tests (Incomplete Sentence Blank, Draw-a-Person, Thematic Apperception Test, Situation Tests, and Three-Minute-Tests) three raters were used to score the subjects' responses. Two graduate students in psychology and one PhD in psychology served as the raters. For each test a training session for variable scoring was involved. During the training sessions the rating approach was practised and discussed over the rating of ten subjects from a police recruit class not involved in this study. Because of the often questioned reliability of ratings of this nature three approachs to measuring reliability were used where possible. These reliability measures are reported in the description of the tests along with the scoring methods. The battery of tests were administered in the order that they are presented below.

Wonderlic Personnel Test, Form B (Wonderlic, 1942)

This is a timed 12 minute test of intellectual development. The test contains items referring to word meanings, arithmetic, proverbs, and noticing missing details.

It has been shown in other research that an established relationship between intelligence scores and marks in police training exist. (For example see Mills, et al., 1965; Mullineaux, 1955; and Dubois and Watson, 1950.) This test has been used to predict police performance in other police departments (see Bass, Karsendiek, Mccullaugh, and Pruitt, 1954). It should be noted that qualities of the ideal policeman as selected by citizens and policemen included terms such as "commonsense", "wise", "capable", "clear thinking", and "intelligent". These qualities should be reflected by scores on this test.

This test was chosen because it was hypothesized that policemen would have to be at a certain level of intelligence so that appropriate police-role selection could occur. It was assumed that an average to above-average level of intellectual development would be required for policemen to be able to understand and properly interpret and enforce laws.

Reliability of the Wonderlic. The reliability of this test as reported in the test manual gives test re-test correlations of r = +.82, +.94 indicating good reliability. Odd-even correlations also resulted in high correlations, r = +.94, again indicating good reliability.

Incomplete Sentence Blank--Adult Form (Rotter and Rafferty, 1955) Plus Addition

This is a projective test which contained 40 sentences stems from the original test plus an additional 17 sentence stems which were added specifically for these selection procedures (see Appendix 4, p. 255). The subject was instructed to: "Complete these sentences to express your real feelings".

A standardized scoring procedure for this test was developed. The test was scored by three raters who were trained in the scoring method. The scoring form is shown in Appendix 5, page 259. Each sentence completion was placed in one of the following categories by each of the three raters: "Harmless Generalization", "Harmful Generalization", "Contribution to Society", "Destructive to Society", "Feelings of Adequacy", "Feelings of Inadequacy", "Liking for Others", "Disliking for Others", "Liking: Nonhuman", "Disliking: Nonhuman", "Others Like Self", "Others Dislike Self", "Self-Serving", "Response Omitted". After categorizing all the responses for a subject the raters gave an overall "Level of Adjustment" score. This scoring approach was designed so that the subject's positive and negative attitudes towards others, towards himself, and towards objects could be ascertained. All of the category scores plus the "Level of Adjustment" score were used to predict the future job performance of the subjects.

This test was selected for a number of reasons. Although it has been infrequently used in previous police research, for this study, the scoring method was structured to indicate functioning in areas of importance for policemen (e.g., "Liking for others", "Disliking for others", etc.). Also, it is a projective test and is less easily "faked good" by police applicants.

It has been previously shown that new poliwemen should have the potential to develop a mature perspective towards citizens and toward the community. To acquire a concept of policing that incorporates roles varying from law-enforcer to public servant necessitates a personality free of unhealty perceptions and attitudes. Results from this test may indicate the perceptions and attitudes of the subjects.

As has been mentioned previously, Anastasi (1968) has pointed out that personality inventories can be altered in the desired direction. Using projective tests will lessen this problem of police applicants presenting their best image and distorting test results.

Reliability of the Incomplete Sentence Blank. Three raters gave an "Overall Rating" score for the sentence completion test attempted by each subject. Three measures of reliability were used to ascertain the consistency of the scoring approachs among the raters. One method used to establish between-rater reliability was the calculation of between-rater correlations for the "Overall Rating" score. These correlations are shown in Table 10 and indicate that a significant positive relationship between raters occurred. With this approach to measuring reliability, however, certain assumptions have been violated. A correlational approach assumes variables are either of the interval type or of the ratio type. While rating these candidates it is most likely that raters operate on an ordinal type of scale which allows for statements such as "this subject is better than that subject". The reduction of the data to an ordinal level diminishes the power of the correlational approach.

Table 10

Means and Correlations of Raters on "Overall Rating" scale for the Incomplete Sentence Blank

Raters:	Rater 1	Rater 2	Rater 3
Mean	2.893	2.676	2.737
Correlation with:			
Rater 1	1.000	.440	.676
Rater 2	.440	1.000	.450
Rater 3	.676	.450	1.000

It should also be pointed out that the correlational approach measures the similarity in the variability of the raters' scores. If two raters are rating consistently close to the same scale points and yet with variability around these points then the correlation will be low (if the points are close together) and the reliability will still be high. With the means, as shown in Table 13, as close as they are, the few low correlations may be more indicative of rater variability within a small range on the scale. These correlations may not indicate that low reliability between raters has occurred.

Lawlis and Lu (1972) have developed a chi-square method of examining the reliability among raters that violates fewer statistical assumptions. In designing their approach they have assumed that every judgment made by a rater has an equal probability of occurring at any point along the number scale. It was also assumed that the judges had no understanding of the rating system and that their judgments were purely random. The statistical technique employed involved comparing the probability of judges agreeing on the rating of a subject with the probability of these agreements occurring by chance. Flexibility was also included in this approach. The authors have stated that the range on the number scale within which an agreement was said to occur could be changed. For example, suppose that judges were required to rate candidates on a five-point scale. The range of agreement might be set at a deviation of one (r=1). In this case a judge scoring five would be considered in agreement with a judge scoring four. This type of flexibility does not violate the basic assumptions of the chi-square approach because the probability of two adjacent categories being scored out of five are considered against the chance probability of this occurrence.

For the five-point scale used in scoring the Incomplete Sentence Blank the number of agreements occurring and the number of expected agreements over three scale interval lengths were calculated. That is, the number of agreements within a scale width of zero (perfect agreement), r=0; the number of agreements within a scale width of one, r=1; and the number of agreements within a scale width of two, r=2; were independently ascertained and compared with the expected agreements for these different interval widths. These results are shown in Table 11.

In scoring this test the raters were allowed to mark at the quarter points along the five-point scale. Because of this fact few perfect agreements occurred among the three raters. When a range of one is considered an agreement, a great deal of reliability among raters is demonstrated. Similarly with a range of two. This scoring procedure is shown to be significantly reliable for scale widths of one and two.

A third method of measuring reliability was also employed. Ebel (1951) proposed an analysis of variance method to compute "intraclass correlation" between raters. It involved obtaining a measure of variance across subjects' scores and a measure of variance across the raters. For these data, the F statistic was, $F = 4.249$, $p<.001$. This demonstrated that rater variance plus error variance was much lower than subject score variance and that a reliable scoring technique was in use.

All three methods of measuing reliablity have shown that the scoring of this test was completed in a reliable fashion.

Table 11

Chi-Square Reliability Measures of the Incomplete Sentence Blank

	Perfect agreement r=0	Range of 1 Considered agreement r=1	Range of 2 Considered agreement r=2
Number of times the 3 raters "agree"	1.00	43.00	57.00
Expected agreements	2.44	14.15	31.72
Chi Square	1.56	75.14	42.06
Probability	not significant	p<.001	p<.001

Draw-a-Person (DAP) (Machover, 1949)

This projective test required the subjects to "draw a person" on a blank sheet of paper (8 1/2" x 11"). This test was chosen to indicate the subjects' self concepts. It may be possible that self images that involve an over-reliance on authority as a means of solving problems, and attitudes that are anti-social in nature can be measured on this test. Similarly, more socially acceptable self-images may be measurable. These measure, may have a strong relationship with a subject's potential to develop "good" concepts of policing. The test also has the advantage of being of the projective type and therefore is less likely to be "faked good".

The drawings were scored by three raters after a training session with a scoring method which was designed for this research. The scoring forms are shown in Appendix 6, page 261. Each drawing was scored on a 1-5 scale in terms of the following categories: "Size", "Placement", "Sheet Position", "Age of Drawn Person", "Sex of Drawn Person", "Errors", "Artistic Quality", "Identity (clothing)", "Line Quality", "Proportion", "Mental Development", "Stance", "Facing Position", "Action", "Partner", "Mouth", "Neck and Chin", "Nose", "Eyes", "Head (hair, ears)", "Facial Expression", "Shoulders", "Hands", "Arms", "Chest", "Waist", "Genital Region", "Hips", "Legs", "Feet", "Ground Line". The raters were required to assess on a 1-5 scale the overall "Level of Adjustment" of the subjects. These variables were used to predict the future job performance of the subjects.

Reliability of the Draw-a-Person. The three raters scored each drawing on the thirty five-point scales. The sums of these scales were calculated to arrive at an "Overall Total" score. The "Overall Total" scores for the three raters were correlated and are shown in Table 12 along with the means for each rater. These correlations are extremely high. Considering this and also that the means for each rater are very similar it can be concluded that this rating approach is highly reliable.

Thematic Apperception Test (TAT) (Murray, 1944)

For this test the subjects were shown slides of Murray's TAT cards. Slides of cards 1, 5, 7, 13, and 14 were presented to the subjects with the following directions.

> I am now going to show you some slides. For each slide I want you to write a story. In your story please include a past, what has gone on before the slide was taken; a present, what is presently happening in the slide; and a future, what will happen. Also in your story include what the people are thinking and feeling. This means your story should contain a past, a present, a future, thoughts, and feelings. Here is the first slide.

Each slide was presented for four minutes. This projective test was scored using a scoring system devised for this project which is shown in Appendix 7, page 267. Each story was scored by having the three raters organize it into the following categories: "Problem set by Other(s)", "Problem set by Self", "Feelings of Adequacy", "Feelings of Inadequacy", "Solution by Other(s)", "Solution by Self", "Good Solution to Problem", "Poor Solution to Problem", "Liking for Other(s)", "Disliking for Other(s)", "Story Incomplete, none". Once the three raters had categorized

Table 12

Means and Correlations of Raters on "Overall Total" Scale for the Draw-a-Person

Raters:	Rater 1	Rater 2	Rater 3
Mean	91.99	99.71	88.35
Correlation with:			
Rater 1	1.000	.923	.887
Rater 2	.923	1.000	.914
Rater 3	.886	.914	1.000

the story in this manner they were required to crystallize and state the fundamental belief or "credo" upon which the story was based. This "credo" was rated on a 1-5 scale according to a "Level of Adjustment".

This test was chosen to indicate the qualites of personality that should affect police performance. Core beliefs shown in this test should have a direct relationship to job performance. Core beliefs may be the basis for the development of concepts of policing. Also, it is these core beliefs which may be the basis for role selection in police work. For example, a core belief such as "Nice guys finish last" as the underlying principle in the projected story of the subject would rate lower than the core belief "A winner never quits".

Reliability of the TAT. The six TAT slides were scored by three raters and given an "Overall Level of Adjustment" rating on a five-point scale. (The scoring followed a training session where the scoring method used was practised on test results of ten individuals, similar to the subjects, but not included in this study.) The three method of measuring reliability that were used for the Incomplete Sentence Blank were also used for the TAT. The correlations among the three raters are shown in Table 13. The correlations range from r =+.143 to r =+.585 with most of the correlations over r =+.30. As was pointed out previously, certain statistical assumptions are violated when this approach is used. (Also, rater variability within an acceptable agreement range may be responsible for the low correlations.) Nevertheless, all of the correlations are in the positive direction and most are significant.

The chi-square approach indicated that the scoring procedure can be accepted as reliable. The chi-square reliability results are shown in Table 14. Rater agreement occurred significantly (p<.001) for all three agreement intervals, r=0, r=1, r=2.

Ebel's analysis of variance method also indicated significant reliability with the F statistic ranging from F =1.983 for card 7 to F =3.364 for card 1 (p <.05 when

Table 13

Correlations between Ratings
on the six TAT Slides

Card 1

Rater:	1	2	3
Rater 1	1.000	.585	.388
Rater 2	.585	1.000	.429
Rater 3	.388	.420	1.000

Card 3

1	2	3
1.000	.328	.289
.328	1.000	.396
.289	.396	1.000

Card 4

Rater:	1	2	3
Rater 1	1.000	.346	.308
Rater 2	.346	1.000	.239
Rater 3	.308	.239	1.000

Card 7

1	2	3
1.000	.143	.428
.143	1.000	.160
.428	.160	1.000

Card 13

Rater:	1	2	3
Rater 1	1.000	.318	.366
Rater 2	.318	1.000	.342
Rater 3	.366	.342	1.000

Card 14

1	2	3
1.000	.424	.347
.424	1.000	.531
.347	.531	1.000

F >1.45). These three methods of measuring reliability indicate that the judges consistently used the same scoring approach.

Table 14

Chi-Square Statistics

for the TAT

	Card 1			Card 3		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (61 Ss)	14.00	50.00	61.00	16.00	51.00	61.00
Expected Agreements	2.44	14.15	31.72	2.44	14.15	31.72
Chi Square	52.62	116.49	56.40	73.28	123.13	56.40
Probability	p<.001	p<.001	p<.001	p<.001	p<.001	p<.001
	Card 4			Card 7		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (61 Ss)	21.00	55.00	61.00	14.00	53.00	61.00
Expected Agreements	2.44	14.15	31.72	2.44	14.15	31.72
Chi Square	139.88	151.53	56.40	52.62	139.96	56.40
Probability	p<.001	p<.001	p<.001	p<.001	p<.001	p<.001
	Card 13			Card 14		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (61 Ss)	16.00	53.00	61.00	18.00	51.00	61.00
Expected Agreements	2.44	14.15	31.72	2.44	14.15	31.72
Chi Square	73.28	139.96	56.40	97.36	123.13	56.40
Probability	p<.001	p<.001	p<.001	p<.001	p<.001	p<.001

Strong Vocational Interest Blank (SVIB) (Strong and Campbell, 1966)

This interest test contains 22 Basic Interest Scales. These scales were derived to show fields of interest towards which the subject is inclined (e.g., "public speaking", "adventure", "social service"). The interest test also contains 57 Occupational Scales. These scales allow for the subject's responses to be compared with the profiles of different occupational groups. For example if the subject responded in a manner similar to experienced policemen, he would receive a high score on the "policeman" scale.

This test was selected because of the established relationships between interest variables and job performance (see Thweatt, 1972; Blum, 1964). It is suggested that interests in the helping professions and leadership roles in association with interests in police work would be of benefit for developing policemen. The relationships between interests and job performance for Canadian policemen will also be clarified in this study.

Reliability of the SVIB. The SVIB has good reliability. It is reported in the test manual (Campbell, 1971) for the Occupational Scales that "test-retest correlations over 30 days average slightly over .90, dropping to about .75 over 20 years for adults and to .55 over 35 years for men first tested at age 16" (p. 21). The author also points out for the Basic Interest Scales that "test-retest correlations are a few points lower than those of the occupational scales" (p. 21). These reliability measures are well within the acceptable range.

Cornell Index Form N2 (Weider, Wolff, Brodman, Mittelmann, Wechsler, 1948)

This test was designed for "...rapid psychiatric and psychosomatic evaluation of large numbers of persons in a variety of situations" (Weider, et al., 1948, p. 2). It can be used to discriminate persons with serious personal and psychosomatic disturbances from the rest of the population. It is correlated with the MMPI (correlating with the neurotic triad, r =+.47, +.60; Weider, et al., 1948).

This test was included as a general screening device to indicate candidates quite different from the normal population. As has been pointed out previously policemen have a higher incidence of ulcer and digestive tract problems (United States Department of Health, cited in Earle, 1973; Los Angeles County Sheriffs Department, 1971). This test should isolate those subjects predisposed to psychosomatic reactions to stressful conditions (which may be part of the job environment of policework). Also because of the higher incidence of suicide within the profession of policing (Heiman, 1975) this test may serve as an indication of those subjects who deal poorly with anxiety and if taken into professions such as police work may be unable to cope.

Reliability of the Cornell Index. Very high reliability estimates for this test are included in the test manual. The Cornell Index Manual states the following. "The reliability coefficient obtained by the Kuder-Richardson formula for one thousand subjects tested at five induction stations is r =+.95" (Weider, Wolff, Brodman, Mittleman, and Wechsler, 1948, p. 3).

Hostility and Direction of Hostility Questionnaire (HDHQ) (Caine, Foulds, and Hope, 1967)

The HDHQ is presented in the form of true-false questions. From fifty questions, two primary measures are obtained: the amount of hostility and the direction of hostility, either intropunitive or extrapunitive. These two primary measures are based on five scales. The outward hostility scales are "urge to act out", "criticism of others", and "projected delusional hostility". The inward hostility scales are "self criticism" and "guilt". These five scales are summed for amount of hostility and combined according to a formula to ascertain direction of hostility. In the manual of the test the authors have stated that "the HDHQ is designed to sample a wide, though not exhaustive, range of possible manifestations of aggression, hostility or punitiveness" (Caine, et al., 1967, p. 5). This test was included in the test battery because one of the essentials for good devloping "good" concepts of policing is the ability to handle coercion effectively. An inappropriate method of hostility expression may prevent the development of this ability. It was also believed that this test would screen out those individuals applying for positions as policemen who have a deeply felt need to control and manipulate others.

Reliability of the HDHQ. The authors have reported that this test battery was administered to a sample of normal persons who were then retested a year later. Test re-test correlations were: amount of hostility, r =+.75, direction of hostility, r =+.51. The authors concluded:

> If we assume that Hostility (amount) has a reliability of .75 and a standard deviation of 6.5, and that Direction has a reliability of .50 and a standard deviation of 5.0, then we may calculate the standard error of measurement as approximately 3.5. (Caine, et al., 1967, p. 14).

These reliability statistics are not of the highest order but are acceptable.

Wechsler-Bellevue Form I Vocabulary (Wechsler, 1946)

For this test the subjects in a group were read the words from the Wechsler Bellvue vocabulary scale one at a time. Each word was spelled for the subjects. The subjects were required to "Write the meaning of the words." The responses were scored according to the Wechsler-Bellvue Form I Manual. This test was given to assess word knowledge and as a measure of general intelligence. Wechsler has reported in a revised edition of this test that the vocabulary score correlates with the overall intelligence score r= +.87 (Wechsler, 1955, p. 17).

The main reason this test was selected was because of the expected relationship between a candidate's intellectual level and marks in police training classes as has been noted in other police departments (see Mills, et al. 1966; Mullineaux, 1955; Dubois and Watson, 1950). Also, word knowledge may also be related to written and oral expressiveness, two qualities essential for good policemen.

Reliability of the Wechsler Bellevue Vocabulary. In the test manual (Wechsler, 1946) no reliability score is given for the vocabulary sub-test alone. However, the following reliability figures are reported for the combined sub-tests scores for the entire Wechsler-Bellevue Intelligence test. "The standard error of measurement for the full scale (IQ's) whose mean and S.D. are respectfully 100.11 and 14.49, is 5.674" (Wechsler, 1946, p. 133). This is an acceptable reliability measure.

California Psychological Inventory (CPI) (Gough, 1956)

The CPI is a true-false personality questionnaire containing 480 items which on scoring are grouped into 18 personality traits (e.g. "Dominance", "Sociability", etc.).

The CPI has been described as follows.

> The inventory is intended primarily for use with "normal" (non-psychiatrically disturbed) subjects. Its scales are addressed principally to personality characteristics important for social living and social interaction. (Gough, 1969, p. 1)

> It may be expected to find most general use in schools, colleges, business and industry, and clinics and counseling agencies whose clientele consists mainly of socially functioning individuals. (Gough, 1969, p. 44)

This test was selected because of the expected relationship between its personality dimensions and job performance. Police recruits, should have average to above average personality development. Scales on the CPI (e.g., "sociability", "social presence", "well-being", "responsibility", "self-control", "tolerence") should reflect a certain degree of stability and potential for a mature perspective in police work. It should be noted that Hogan (1971) found that certain CPI variables were predictive of later police classroom performance.

Reliability of the CPI. The test manual indicates good reliability measures for the 18 scales. For example, when a group of high school males were given the test, the test-retest correlations ranged from r =+.38 to r =+.75. These correlations may have been slightly lower than expected because of the changing and maturing tendencies of these young subjects. These reliability indicators are acceptable.

Situation Tests and Three-Minute-Tests (3MT)

(Individual tests designed for this research)

These tests were designed to measure the policeman's ability to function in a stressful situation. Verbalizing skills, courtesy, and maturity under stress are variables that should separate different levels of police job performance. For these tests the subjects individually were called upon to perform a number of tasks. These tasks were completed in the presence of three people, one actor and two experimenters. The entire proceedings were recorded on video-tape. The information given the subjects for the different tasks are presented below. The experimenter, actor, and subjects are signified with the abbreviations E, A, and S.

E: "Run to that table and you will receive instructions there."

E: "Put these pieces together to make a whole, one unit." (The subject then proceeded for up to 3 minutes to attempt to assemble the O'Connor Wiggly Block test, O'Connor, 1951.)

E: "Please stop. Would you step over here please? I have some situations I want you to deal with. Here is the first one."

(Lawnchair)

"I want you to imagine that you

have taken your lawnchair down town to watch the Klondike Days Parade. In the middle of the parade you leave your chair to use the washroom facilities. When you return someone is sitting in your chair. Here is your chair and here is the person sitting in it." (The experimenter points to a chair with the actor sitting in it.) "Solve that problem."

A: (Sitting on the edge of the chair and looking forward.) "Look at that float over there. This is the nicest spot I've had in years. Sure lucky I found this chair with a perfect view."

S: .....

A: "Your chair? How do I know it's your chair?"

S: .....

A: "These chairs were put out by the city. I saw them putting some at Sir Winston Churchill Square and they're the same as these."

S: .....

A: "Okay, you can have the chair. I'm tired of sitting anyway."

S: .....

E: "Thank-you. Now would you step over here again please?"

(Directions)

"I want you to imagine that you are standing in front of the legislative building and this gentleman (pointing to the actor) will ask you a quesion."

A: "Excuse me, I'm from out of town and I want to get to city hall. Can you give me directions?"

S: .....

A: "Thank-you."

E: "Thank-you. Now I have another situation for you."

(Lifeguard)

"I want you to imagine that you are a lifeguard and that this individual (points to the actor who begins to run on the spot and then runs forward past the subject) is running on the pool deck which is against the rules. Solve that problem."

S: .....

A: "Who me? What's wrong with running?"

S: .....

A: "I'm a good runner. I've got a few trophies for running. I'm not going to fall."

S: .....

A: "I guess I better stick to the rules."

S: .....

E: "Thank-you. Now would you step over here please, and stand facing me?"

(3MT-People)

"I would like you to speak for three minutes on the question: 'What are people really like?' The time you will have remaining will be shown on this clock." (Points to a clock with a large dial.)

S: .....

E: "Thank-you. Now I have another situation for you."

(Dog)

"I want you to imagine that late at night for the past two weeks you have been disturbed by a dog barking on the back porch of your neighbor. You are going to complain about the noise to the dog's owner. You are now at the door of the owner and are knocking. Simulate the knock by knocking on this table and then begin."

S: (Knocks.)

A: "Yes?"

S: .....

A: "My dog? My dog never barks. It's a very happy dog."

S: .....

A: (A relevant answer maintaining continually the same attitude towards the subject.)

S: .....

A: "My dog barks only when it's teased. Are you teasing my dog? If you are, stop teasing my dog and he'll stop barking."

S: .....

A: "Maybe it is my dog. I'll take him to the vet and get something to quiet him down."

S: .....

E: "Thank-you. Now would you stand here please?"

(3MT-World)

"I would like you to speak for three minutes on 'What's wrong with the world?' The time you will have remaining will be shown on the clock. (Points to a clock with a large dial.) Begin now."

S: .....

E: "Thank you. That's the end of the testing. I want to thank you for cooperating and for enduring this slightly stressful procedure. I ask you not to discuss this part of the testing with your classmates until everyone has been tested. It would be fairer for everyone if all begin with the same knowledge of what is to take place. Thank you again."

The last two tasks (Dog barking, 3MT-World) were added for class 38 and were not given to class 37. This addition should have no diverse effect on the standardization of administration because these changes came at the end of the entire test battery and therefore could not alter the response set of the subjects for following tests.

The situation tests and 3MT tests were scored with a procedure designed specifically for this research. The subjects' behaviors were assessed with the use of videotapes by the three raters with scoring methods designed for the situation tests and the 3MT tests. These rating forms are shown in Appendix 8, page 270 and 271. The situation test scoring form contained the following categories which were scored on a 1-5 scale: "Approach to Situation", "Clarity of Verbal Request or Response", "Consideration Given to Other's Comments", "Effectiveness of Solution to Problem", "Confidence Shown in Handling Situation", "Interpersonal Skills". Once the subjects' behaviors were assessed in these categories an overall rating based on Level of Development was scored on a five point scale. The score on this scale was used to predict future performance as a policeman.

The Three-Minute-Test rating scale contains the following categories which also were scored on a 1-5 scale: "Introduction", "Ideas", "Clarity", "Voice", "Organization", "Reasoning", "Closure", "Poise", "Quality of Content", "Listener's Reaction". Once these scales were scored by the three raters the subjects were rated on "Level of Development" (five point scale). The score on this final scale was used to predict the future performance as a policemen.

Reliability of the Situation Tests and the Three-Minute-Tests. The three raters gave an "Overall Rating" (Level of Development) score to each subject for each of the situation tests and three-minute-tests attempted. Three measures of reliability were applied to the data to ascertain the consistency and similarity of scoring approach by the raters. Table 15 shows the correlations between raters for these six tests.

Table 15

Correlations between Ratings on the Situation Tests and Three-Minute-Tests

	Lawnchair			Directions		
Rater	1	2	3	1	2	3
Rater 1	1.000	.608	.668	1.000	.744	.675
Rater 2	.608	1.000	.458	.744	1.000	.569
Rater 3	.668	.458	1.000	.675	.569	1.000

	Lifeguard			3MT-People		
Rater	1	2	3	1	2	3
Rater 1	1.000	.682	.469	1.000	.484	.635
Rater 2	.682	1.000	.442	.484	1.000	.464
Rater 3	.469	.442	1.000	.635	.464	1.000

	Dog			3MT-World		
Rater	1	2	3	1	2	3
Rater 1	1.000	.591	.252	1.000	.771	.669
Rater 2	.591	1.000	.248	.771	1.000	.518
Rater 3	.252	.248	1.000	.669	.518	1.000

The correlations reported on these tests indicate the following order when considering which tests are the most reliable:

i) Directions (lowest r = +.569)
ii) 3MT-World (lowest r = +.518)
iii) 3MT-People (lowest r = +.464)
iv) Lawnchair (lowest r = +.458

v) Lifeguard (lowest r = +.442)

vi) Dog (lowest r = +.248)

The first three tests, which are the most reliable, primarily involved the subject verbally answering a question with little or no exchange between the actor or experimenter. The last three tests involved situations where the actor had a larger role in communication and the situations contained stressful and mildly hostile elements. From these correlation measures, the "Dog" situation would not be considered reliable enough to include in any furthur analysis. (The disadvantages of using correlations as a measure of reliability have been pointed out previously. Better reliability measures, which are presented below, suggest that this test is acceptably reliable.)

The chi-square approach of measuring reliability indicated that all six of these tests were scored in a reliable manner. The results of the chi-square analysis are shown in Table 16. These results indicate that when a rater agreement is considered at r=1 or r=2 (that is when the range of agreement is either an interval of one or an interval of two) the results are highly reliable. Because the raters were allowed to rate at quarter points along the scale few perfect agreements (r=0) occurred.

The analysis of variance approach to reliability also indicated that the scoring approach was a reliable one with

Table 16

Chi-Square Statistics

for the Situation Tests and The Three-Minute-Tests

	Lawnchair			Directions		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (61 Ss)	3.00	55.00	61.00	1.00	58.00	61.00
Expected Agreements	2.44	14.15	31.72	2.44	14.15	31.72
Chi Square	0.03	155.82	55.48	1.52	179.57	55.48
Probability	p<.400	p<.001	p<.001	p<.001	p<.001	p<.001

	Lifegaurd			People		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (61 Ss)	2.00	42.00	61.00	0.00	44.00	60.00
Expected Agreements	2.44	14.15	31.72	2.44	14.15	31.72
Chi Square	0.34	72.37	55.48	3.57	83.15	51.70
Probability	p<.400	p<.001	p<.001	p<.001	p<.001	p<.001

	Dog			3MT-World		
	r=0	r=1	r=2	r=0	r=1	r=2
3 Raters Agree (28 Ss)	0.00	22.00	28.00	0.00	25.00	28.00
Expected Agreements	1.12	6.50	14.50	1.12	6.50	14.50
Chi Square	2.36	46.57	25.96	2.36	66.69	25.96
Probability	p<.400	p<.001	p<.001	p<.001	p<.001	p<.001

F statistics as follows:

i)	Lawnchair	F = 4.751
ii)	Directions	F = 6.149
iii)	Lifeguard	F = 3.676
iv)	3MT-People	F = 4.056
v)	Dog	F = 2.482
vi)	3MT-World	F = 5.842

All of these F scores are significant ($p<.05$). The scoring procedure for these tests can be assumed to have been reliable.

Experimental Design

The 61 subjects were divided into three groups according to their actual later job performance after being with the police force for two years. Membership in these groups; high performers, average performers, and low performers; was determined by having each subject rank-ordered in his squad by his sergeant and then having the sergeants rate the subject's overall performance.

These two measures were obtained in the following way. The 22 sergeants involved were given an instruction sheet and a number of scoring forms equal to the number of men in their squad (see Appendix 9, p. 272). The instruction sheet was read to each sergeant and then the sergeant assessed the squad members on the two variables (rank-order first and then overall performance). If questions from the sergeants arose concerning what constituted "overall performance" it was explained as follows.

> If an emergency of an unknown nature occurred for which you were to send your best police constable, the man you would send should be ranked first. If that man was not available, the man you would send next would be ranked second and so on.

Each sergeant was required to have had the subject being rated in his squad for at least two months. Also, if the sergeant stated that a new squad member would be difficult to rate because he did not know him well enough, then this subject was rated by the previous sergeant who would have known the candidate for a longer period of time.

For the overall performance ratings, score distributions tended to highly favor the "Above Average" side of the rating continuum. It should be noted that for this measure tied scores between or among squad members were allowed. Therefore, it was possible for a sergeant to give the top four squad members a score of eight in terms of overall performance and the next four a score of seven. This resulted in a narrower distribution of scores than the rank-order measures which were spread across the continuum (tied ranks not being allowed for this measure). It was arbitrarily decided to use the rank-order measures for the criterion because the distribution resulting from these scores would be better suited for the sorting type of statistical analysis that will be used in the study.

Reliability of the performance measures. The basic research was completed (including the performance ratings of March-April 1976) and the statistical calulations were made. Two performance appraisals periods have since followed and a large amount of reliability information has been gathered. The three performance appraisals were carried out at the following times. In March-April 1976 the sergeants involved rank-ordered the men in their squad and then rated their overall performance on a nine point scale. In March-April 1977 a similar rating and ranking occurred. In January-February 1978 the sergeant, the senior constable in the squad, and the staff-sergeant ranked and rated the performance of the subjects. The staff sergeants, because there was a question of how familiar they were with the job performance of the subjects, also indicated "how well" they knew the subject. (See Appendix 10, page 275, for a further elaboration of the third appraisal period.)

Table 17 shows the correlations among the ranking measures for the five raters involved. The correlation between sergeants for 1976-1977, as shown in this table, was r=+.5134; for 1977-1978, r=+.6715. As the time between ranking increased the correlation became lower, 1976-1978, r=+.4768. These total results indicate that this ranking measure was reliable over the three year time period.

In 1978 the senior constables in the squad also ranked the squad. Their ranking correlated with the 1978

Table 17

Correlations Among Ranking Measures

	Serg. Ranking 1976	Serg. Ranking 1977	Serg. Ranking 1978	Const. Ranking 1978	Staff Serg. Ranking 1978
Serg. Ranking 1976	1.000	.5134	.4768	.2988	.3501
Serg. Ranking 1977	.5134	1.000	.6715	.4221	.4146
Serg. Ranking 1978	.4768	.6715	1.000	.5033	.4920
Const. Ranking 1978	.2988	.4241	.5033	1.000	.5824
Staff Sergeant Ranking 1978	.3501	.4146	.4920	.5824	1.000

sergeants' ranking, r=+.5033. The staff-sergeants' ranking in 1978 correlated with the 1978 sergeants' ranking, r=+.4920. All of these correlations are also significant and indicate a reliable performance appraisal.

The staff sergeants (n=2) in 1978 were also asked to rank and rate the subjects in the study only if they knew them well enough. After this decision, before ranking and

Table 18

Comparison of Correlation and Partial Correlation of Ranks

	Correlation of Staff Sergeant Ranking 1978	Partial Correlation of Staff Sergeant Ranking 1978
Constable Ranking 1978	.5824	.4229
Sergeant Ranking 1978	.4920	.3794

rating the subjects, they indicated "know well", or "know very well" regarding their rating of the subjects. When this variance is eliminated from the correlational pairing a slightly lower correlation results as shown in Table 18. These results demonstrate that when the variance for "how well known" is subtracted the correlations are still reliable.

The correlations for the rating measures given on the nine point scale were also shown to be reliable. Table 19 indicates that over a one year period of time the correlations for the rating pairings (sergeant-sergeant) were; 1976-1977, r=+.5030; 1977-1978, r=+.4837. Over the two year period the correlations were slightly lower, 1976-

Table 19

Correlations Among Rating Measures

	Serg. Ratings 1976	Serg. Ratings 1977	Serg. Ratings 1978	Const. Ratings 1978	Staff Serg. Ratings 1978
Serg. Ratings 1976	1.000	.5033	.3220	.1663	.2663
Serg. Ratings 1977	.5030	1.000	.4837	.2533	.3726
Serg. Ratings 1978	.3220	.4837	1.000	.2577	.4706
Const. Ratings 1978	.1663	.2533	.2577	1.000	.4799
Staff Sergeant Ratings 1978	.2663	.3726	.4706	.4799	1.000

1978, r=+.3220. In 1978 the sergeants' and constables' ratings correlated r=+.2577, sergeants and staff-sergeats, r=+.4706. The low correlation between constables and sergeants may reflect differing definitions of what comprises "very good" performance. In general the rating measures are considered reliable.

For all performance appraisals two measures of overall performance were required of each rater. This allowed for an intra-rater reliability measure, somewhat similar to the split-half technique used for psychological tests. The intra-rater reliability measures were as follows: 1976 sergeants, r=+.7774; 1977 sergeants, r=+.6403; 1978 sergeants r=+.5809; 1978 staff-sergeants, r=+.8681; 1978 constables, r=+.6227. These measures indicate within the raters a consistent definition of "overall performance".

Another form of measuring reliability was also considered. During the second assessment period, immediately following the obtaining of the ranking and rating measures, the sergeants were required to rate three written incidents of job performance on the same nine point scale of overall performance (See Appendix 11, page 284). The chi-square approach to obtaining reliability was used and results were as shown in Table 20. Significant reliability was obtained for all three written incidents. This significant level of agreement among the sergeants indicates that their judgment of job performance as presented in written form was basically in agreement. That is, the sergeants tended to agree on what constituted good and poor job performance. These results may imply that similar standards of job performance are being considered when sergeants are rating constables. This type of agreement among sergeants indicates that a standardized and reliable approach to measuring written descriptions of job performance occurred. For the interests of the reader the

Table 20

Chi-Square Reliability Measures
of the Three Incidents

	Intoxication Incident r=0	Speeding Incident r=0	Maritial Dispute Incident r=0
Number of times the raters "agree"	9.00	10.00	9.00
Expected agreements	2.88	2.88	2.88
Chi Square	14.91	20.18	14.91
Probability	p<.001	p<.001	p<.001

average ratings and standard deviations of the incidents are presented in Appendix 11, page 287.

A third estimate of the reliability of the scoring approach of the sergeants was also used. Following the rating of the three incidents, the sergeants for the second assessment period (n=21) rated twenty attributes of job behavior "in terms of their importance for the _overall performance_ of a policeman". These attributes were also rank-ordered. The rating sheet containing the attributes and the method of administration is shown in Appendix 12, page 288. Results from these investigations show that policemen reliably rank-order important attributes required for police work. The five attributes that were seen as the most important are shown in Table 21.

It is clearly shown in Table 21 that the five attributes ranked highest by the sergeants were weighted in terms of "overall importance" in a reliable fashion, all 21 sergeants agreeing (beyond chance effects) upon the relative importance of these attributes. The high weightings given these attributes suggest that they are among the important qualities that the sergeants were considering when they were rating overall performance.

The three different approachs to measuring reliability indicate that the method of performance assessment used by the sergeants was reliable. The correlations among raters were consistently significant. Also, the concepts of good and poor job performance held by the different sergeants were demonstrated to be similar. Incidents of good and poor job performance were consistently rated in the same manner by the sergeants. The sergeants significantly agreed upon the important attributes of job performance. It can be concluded that the rating and rank-ordering approach to performance assessment was carried out in a reliable manner.

The criterion selected (rank-order score) is considered to be the best possible. The value of the study is dependent upon this measure being representative of job

Table 21

Most Importat Attributes of Policework as Selected by Police Sergeants

Attribute	Placement in top 5	Average Rating	Standard Error	Range	Chi Square*
(1) Commonsense	100.0%	8.90	.04	1	57.10
(2) Honesty	81.0%	8.81	.09	1	57.10
(3) Self-Control	71.4%	8.62	.15	2	28.00
(4) Fairness	57.1%	8.09	.18	3	14.59
(5) Alertness	47.6%	8.48	.16	2	28.00

*p<.01

performance.

The following analyses were conducted. Means, standard deviations, and univariate F statistics for the three groups on all of the variables were found. Multiple discriminant analyses were applied to each test with four or more variables. (This statistical technique loses its applicability when an extremely small number of variables are included in the analysis.) The results of each analysis will be an equation (or equations) which will provide the weighting coefficients for the subset of variables that will best discriminate among the three groups. The best discriminating variables for each test (with four or more variables) will then be combined with variables previously not included in a multiple discriminant analysis and entered into a large multiple discriminant analysis for the entire test battery. This approach will demonstrate the value of each test independently and the value of the entire test battery in statistically discriminating among the three groups. Also indicated will be the order of the best discriminating variables.

CHAPTER III

RESULTS

Wonderlic Form B

Three variables were scored on this test: the number of wrong answers, the number of correct answers, and the number of questions attempted. Only two of these variables (number wrong, number correct) were entered into the statistical analyses so that statistical independence among the variables could be maintained. The means, standard deviations, and F ratios for the two variables are shown in Table 22. No significant differences occurred across the three groups for these two test variables. The test manual indicates a score of 22 to 23 is at the 39.2 percentile, while scores 26 to 27 are at the 61.6 percentile. (These percentiles were based upon the scores of males ages 18 to 30 who had completed three to four years of high school.) These results show that these police recruits fall within the average range of intellectual functioning.

Table 22

Group Means (Standard Deviations) and
F Statistics for the Wonderlic

Wonderlic Variable	High Level Performers	Average Level Performers	Low Level Performers	F
Number Wrong on Wonderlic	4.97 (4.47)	7.22 (2.95)	5.37 (2.83)	2.60
Number Correct on Wonderlic	26.48 (9.67)	24.91 (3.79)	22.21 (6.21)	.68

Incomplete Sentence Blank

The three raters' scores on each scale of the Incomplete Sentence Blank were summed. The means of these sums for the three performance groups are shown in Table 23.

When the 15 variables used in this test are considered the "Response Omitted" variable appears as the best discriminator, $F = 3.91$, $p<.05$. The high level performers omitted more responses then did the other two groups. The second best discriminating variable on this test was the "Feelings of Inadequacy" measure. The high level performers responded to the sentence stems with fewer feelings of inadequacy. This difference however, was not a signficant one ($p<.25$). All of the other variables also did not

Table 23

Group Means (Standard Deviations) and F Statistics for the Incomplete Sentence Blank

Sentence Completion Category	High Level Performers	Average Level Performers	Low Level Performers	F
Harmless Generalization	27.15 (15.28)	34.91 (16.25)	35.10 (17.27)	1.57
Harmful Generalization	1.85 (1.57)	3.18 (3.10)	2.90 (3.89)	1.13
Contribution to Society	6.75 (6.01)	7.91 (5.03)	8.26 (5.70)	.40
Destructive to Society	00.00 (.3244)	0.36 (0.85)	0.15 (0.50)	1.87
Feelings of Adequacy	24.30 (10.55)	24.18 (10.21)	23.94 (8.54)	.01
Feelings of Inadequacy	22.10 (11.38)	28.59 (99.91)	26.47 (10.67)	2.00
Liking for Others	12.60 (7.88)	15.50 (8.34)	12.79 (97.55)	.88
Disliking for Others	5.30 (4.14)	5.91 (4.77)	5.57 (3.74)	.11
Liking: Nonhuman	12.05 (6.81)	10.63 (6.20)	10.53 (4.87)	.40
Disliking: Nonhuman	4.70 (3.11)	5.95 (3.81)	6.21 (4.06)	.96
Others like Self	1.25 (1.83)	1.72 (4.47)	1.11 (2.24)	.22
Others Dislike Self	.20 (.70)	.32 (1.52)	.05 (.40)	.34
Self Serving	12.70 (5.32)	12.00 (5.72)	13.84 (9.46)	.36
Response Omitted	30.60 (28.69)	11.95 (18.07)	13.84 (22.66)	3.91*
Overall Level of Adjustment	8.44 (2.72)	8.26 (1.99)	8.22 (1.83)	.05

*$p<.05$

significantly discriminate among the three performance groups.

A stepwise multiple discriminant analysis for these 15 variables was applied to the data. It was found that maximum separation among the three groups could be achieved with four of the 15 variables. These variables were "Response Omitted", "Destructive to Society", "Liking-nonhuman", and "Disliking-nonhuman". This finding indicates that the variance accounted for by these four variables can be combined along two discriminant functions to maximally differentiate among these three groups. For these subjects the remaining eleven variables did not contribute any unique discriminating power not already contained in the previously mentioned variables. Two discriminant functions were derived which resulted in a correct classification of 47.54% of the three performance groups. The classifications as they occurred are shown in Table 24.

The F ratios for the discriminant functions developed are as follows:

F across all groups, $F = 3.47$, $p < .05$
Group 1 and Group 2, $F = 3.76$, $p < .01$
Group 1 and Group 3, $F = 2.66$, $p < .05$
Group 2 and Group 3, $F = 0.33$, (not significant).

The above statistics indicate that this test discriminates

Table 24

Prediction Results of
Incomplete Sentence Blank

Actual Group Membership	Number of Cases	Predicted Group Membership		
		Group 1	Group 2	Group 3
Group 1 High Level	20	13 65.0%	1 5.0%	6 30.0%
Group 2 Average Level	22	5 22.7%	7 31.8%	10 45.5%
Group 3 Low Level	19	5 26.3%	5 26.3%	9 47.4%

Percent of Cases Correctly Classified: 47.54%

between groups one and two, and between groups one and three, but not between groups two and three. It follows that this test is best used in selecting high level policemen from among the average and above average candidates. It is not a good measure to use when choosing which candidates from a low level group should not be hired.

Draw-a-Person

The average raters' scores for each scale devised for the Draw-a-Person test across the three performance groups are shown in Table 25 along with the F statistics. As shown in the table none of the variables significantly

Table 25

Group Means (Standard Deviations) and
F Statistics for the Draw-a-Person Test

Draw-a-Person Category	High Level Performers	Average Level Performers	Low Level Performers	F
Size	3.09 (1.54)	3.05 (1.25)	3.10 (1.41)	0.01
Placement	3.50 (0.88)	3.53 (0.51)	3.60 (0.93)	0.06
Sheet Position	3.76 (0.90)	4.00 (0.00)	3.56 (1.07)	0.80
Age	3.63 (1.02)	3.89 (0.42)	3.53 (1.09)	0.88
Sex	3.60 (1.02)	3.72 (0.66)	3.76 (0.92)	0.21
Errors	3.62 (1.02)	3.83 (0.33)	3.56 (0.92)	0.68
Artistic Quality	3.58 (0.97)	3.80 (0.39)	3.32 (0.96)	1.80
Identity	3.11 (1.43)	3.17 (1.02)	2.75 (1.30)	0.68
Line Quality	3.54 (0.95)	3.78 (0.52)	3.30 (0.95)	1.78
Proportion	3.54 (0.97)	3.66 (0.53)	3.40 (1.02)	0.46
Mental Development	3.72 (1.01)	3.92 (0.49)	3.48 (0.93)	1.40
Stance	2.31 (1.07)	2.80 (1.07)	2.26 (1.19)	1.54
Facing Position	3.72 (1.49)	3.92 (1.26)	3.84 (1.51)	0.11
Action	1.98 (0.59)	2.27 (0.65)	1.97 (0.64)	1.54
Mouth	3.25 (1.12)	3.22 (0.96)	3.02 (1.00)	0.30

Table 25 (continued)

Group Means (Standard Deviations) and
F Statistics for the Draw-a-Person Test

Draw-a-Person Category	High Level Performers	Average Level Performers	Low Level Performers	F
Neck and Chin	3.53 (1.10)	3.63 (0.67)	3.49 (1.15)	0.10
Nose	2.93 (0.92)	3.21 (0.70)	2.94 (1.02)	0.69
Eyes	2.79 (1.01)	2.96 (0.74)	2.88 (1.01)	0.18
Head	3.12 (1.04)	3.33 (0.55)	3.17 (0.94)	0.39
Facial Expression	3.21 (0.96)	3.36 (0.77)	3.25 (0.99)	0.16
Shoulders	2.74 (1.27)	3.16 (1.13)	2.79 (1.34)	0.72
Hands	1.68 (0.88)	1.97 (0.83)	1.50 (0.78)	1.69
Arms	2.48 (1.16)	2.75 (1,14)	1.90 (1.03)	3.02
Chest	2.99 (1.46)	3.07 (1.17)	2.55 (1.36)	0.87
Waist	2.97 (1.38)	2.98 (1.29)	2.52 (1.50)	0.71
Genital Area	2.54 (1.33)	2.71 (1.28)	2.26 (1.39)	0.59
Hips	2.68 (1.40)	2.86 (1.25)	2.30 (1.32)	0.93
Legs	2.57 (1.49)	2.40 (1.22)	2.11 (1.26)	0.58
Feet	1.95 (1.06)	2.03 (1.09)	1.68 (0.95)	0.60
Ground Line	2.18 (0.90)	2.39 (0.97)	1.85 (0.64)	2.05
Total Score	90.30 (25.16)	95.38 (13.67)	85.65 (24.23)	1.06

discriminated among the three groups. However, the variable "Arms of Drawn Person" approached significance (p<.10). The low performance group, as would be expected, scored lower on this variable than did the other two groups. The variables "Ground Line" and "Artistic Quality" were the next two best discriminating variables when considered independently (p<.25).

Although none of the variables (when taken alone) significantly discriminated among the groups, the multiple discriminant analysis method (which considers the best interactions among the variables) did demonstrate a good separation of the three groups. The variables included in this analysis were: "Placement of Drawn Person", "Sheet Position", "Sex of Drawn Person", "Identity of Drawn Person", "Line Quality of Drawn Person", "Stance of Drawn Person", "Neck and Chin of Drawn Person", "Nose of Drawn Person", "Facial Expression of Drawn Person", "Shoulders of Drawn Person", "Hands of Drawn Person", "Arms of Drawn Person", "Waist of Drawn Person", "Legs of Drawn Person", and "Ground Line". These variables when weighted and linearly combined on two discriminant functions resulted in a maximal separation of the three groups for this test. The prediction results of this one test to the on-the-job performance by the subjects are shown in Table 26.

The F ratios for the discriminant functions developed

Table 26

Prediction Results of the Draw-a-Person Test

Actual Group Membership	Number of Cases	Predicted Group Membership		
		Group 1	Group 2	Group 3
Group 1 High Level	20	13 65.0%	4 20.0%	3 15.0%
Group 2 Average Level	22	2 9.1%	19 86.4%	1 4.5%
Group 3 Low Level	19	1 5.3%	1 5.3%	17 89.5%

Percent of Cases Correctly Classified: 80.33%

are as follows:

F across all groups, F =1.90, p<.01

Group 1 and Group 2, F =1.53, p<.25

Group 1 and Group 3, F =1.44, p<.25

Group 2 and Group 3, F =3.06, p<.005.

These results indicate that the Draw-a-Person could only discriminate significantly between groups two and three. Its best use would be in discriminating among the poorer candidates.

Thematic Apperception Test (TAT)

The three raters scores (on the five point scales) for the overall ratings of the TAT slides were summed and statistical analyses were applied. Means, standard deviations and F statistics for each TAT slide are shown in Table 27. No significant differences occurred for any of the TAT variables. The trend as indicated by TAT Card 1 and TAT Card 7 appeared to be in the opposite direction expected. Those subjects who were scored as having a high "Overall Level of Adjustment" tended to be placed in the Average Level and Low Level performance groups. Because the differences appeared to be in the wrong direction which may have been a function of chance effect, the discriminant analysis statistical procedure was not applied to these TAT data.

Strong Vocational Interest Blank (SVIB)

The means, standard deviations, and univariate F ratios for the variables of the SVIB are shown in Table 28. These results indicate that none of the variables on the SVIB, when considered independently, significantly discriminate among the groups. The best discriminating variables were "Art", "Dentist", "Physician", and "Biologist" on which the high performing group tended to score lower than the other

Table 27

Group Means (Standard Deviations) and F statistics for the TAT

TAT Card	High Level Performers	Average Level Performers	Low Level Performers	F
TAT Card 1	6.95 (2.95)	8.13 (2.20)	7.79 (2.01)	1.45
TAT Card 3	6.57 (2.17)	7.22 (1.83)	7.26 (1.62)	.80
TAT Card 4	7.12 (2.26)	7.48 (1.16)	7.47 (1.58)	.31
TAT Card 7	7.32 (2.11)	8.17 (1.72)	8.21 (1.51)	1.56
TAT Card 13	7.32 (2.56)	7.65 (1.53)	6.84 (2.12)	.79
TAT Card 14	6.90 (2.40)	6.83 (1.85)	6.63 (1.42)	.10
TAT Total Score	41.79 (11.72)	45.22 (6.95)	44.26 (5.78)	.89

groups.

When the multiple discriminant analysis was applied to the data derived from the SVIB near perfect group prediction occurred as shown in Table 29. This discrimination results because the unique discriminating power of each variable was considered in its relationship to the other variables across two discriminant functions.

The F ratios for the discriminant functions developed

Table 28

Means (Standard Deviations) and F Statistics for the Strong Vocational Interest Blank

Interest Area	High Level Performers	Average Level Performers	Low Level Performers	F
Public Speaking	55.45 (15.51)	56.59 (10.11)	53.74 (7.07)	0.31
Law/ Politics	56.85 (15.23)	56.59 (10.11)	53.74 (7.07)	1.01
Business Management	52.40 (14.29)	53.68 (9.22)	50.37 (8.63)	0.47
Sales	51.25 (15.10)	52.41 (7.87)	52.48 (7.94)	0.12
Merchandising	50.85 (14.51)	52.32 (9.04)	49.63 (6.02)	0.34
Office Practices	49.15 (13.37)	52.55 (8.93)	50.68 (8.95)	0.54
Technical Supervision	49.25 (13.66)	50.50 (9.97)	47.00 (10.47)	0.48
Mathematics	45.65 (14.10)	47.09 (10.17)	46.32 (9.95)	0.08
Science	45.65 (13.11)	48.09 (9.39)	48.53 (8.32)	0.44
Mechanical	46.80 (14.87)	50.91 (9.92)	47.26 (7.80)	0.84
Nature	45.35 (13.89)	49.05 (6.93)	47.21 (7.64)	0.73
Adventure	60.05 (15.22)	63.14 (8.03)	64.95 (8.16)	1.00
Recreational Leadership	55.50 (14.52)	59.09 (9.08)	58.84 (7.07)	0.71
Medical Service	53.05 (13.52)	53.96 (9.28)	55.21 (6.77)	0.22
Social Service	58.45 (15.88)	63.77 (7.18)	61.58 (9.83)	1.14

Table 28 (continued)

Means (Standard Deviations) and F Statistics for the Strong Vocational Interest Blank

Interest Area	High Level Performers	Average Level Performers	Low Level Performers	F
Religious Activities	49.15 (14.65)	48.59 (7.36)	50.79 (8.35)	0.23
Teaching	52.65 (14.21)	54.96 (7.26)	56.05 (6.59)	0.60
Music	45.70 (14.14)	40.00 (10.32)	43.16 (8.57)	0.95
Art	45.35 (12.82)	52.23 (8.01)	47.11 (6.37)	3.01[1]
Writing	51.45 (14.17)	51.18 (8.91)	50.74 (7.08)	0.02
Dentist	19.35 (8.41)	23.41 (9.09)	24.74 (9.17)	1.97
Osteopath	26.70 (9.13)	28.36 (8.71)	30.74 (7.23)	1.13
Veterinarian	27.80 (11.92)	29.41 (7.33)	30.11 (7.74)	0.33
Physician	21.30 (8.96)	21.55 (11.21)	26.47 (8.04)	1.82
Psychiatrist	21.60 (10.75)	22.14 (9.93)	23.05 (11.20)	0.09
Psychologist	19.45 (11.31)	18.45 (8.08)	20.95 (6.38)	0.41
Biologist	16.10 (11.78)	17.82 (9.16)	21.58 (8.18)	1.58
Architect	15.10 (10.03)	18.36 (9.70)	17.32 (7.65)	0.67
Mathematician	10.00 (9.54)	11.82 (8.88)	13.16 (7.74)	0.64
Physicist	9.20 (9.42)	9.87 (10.90)	11.47 (10.00)	0.26
Chemist	15.60 (12.70)	15.22 (12.98)	18.05 (12.68)	0.29
Engineer	17.20 (9.74)	18.09 (13.05)	18.16 (9.91)	0.05
Production Manager	30.70 (8.61)	31.18 (11.07)	28.58 (9.20)	0.40
Army Officer	30.45 (11.77)	29.36 (13.38)	28.21 (9.38)	0.18

---

[1] This is the largest F ratio.

Table 28 (continued)

Means (Standard Deviations) and F Statistics for the Strong Vocational Interest Blank

Interest Area	High Level Performers	Average Level Performers	Low Level Performers	F
Air Force Officer	34.40 (11.05)	35.14 (12.15)	33.95 (7.44)	0.18
Carpenter	20.60 (11.48)	25.36 (13.77)	24.32 (11.16)	0.86
Forest Service Man	26.05 (13.33)	31.36 (11.55)	31.79 (9.32)	1.54
Farmer	27.70 (10.72)	31.18 (8.65)	30.74 (9.81)	0.77
Math-Science Teacher	25.25 (9.36)	29.91 (9.15)	28.58 (10.80)	1.25
Printer	31.25 (10.71)	35.00 (9.30)	32.74 (10.28)	0.74
Policeman	37.05 (11.59)	38.68 (11.00)	39.53 (9.26)	0.27
Personnel Director	30.05 (11.79)	33.09 (12.53)	27.68 (12.18)	1.02
Public Administrator	37.95 (13.44)	39.77 (11.75)	34.16 (11.89)	1.07
Rehabilitation Counselor	35.40 (12.53)	35.73 (12.14)	33.58 (11.05)	0.19
YMCA Staff Member	36.40 (11.56)	35.73 (11.31)	33.58 (9.73)	0.43
Social Worker	33.65 (12.83)	34.95 (15.37)	32.53 (13.23)	0.16
Social Science Teacher	34.00 (11.56)	38.09 (12.15)	36.68 (8.39)	0.75
School Superintendent	20.20 (11.67)	20.00 (12.78)	18.84 (8.35)	0.08
Minister	15.45 (10.95)	19.46 (13.92)	16.21 (7.60)	0.75
Librarian	21.25 (10.73)	25.09 (11.65)	23.79 (8.01)	0.74
Artist	19.65 (7.61)	21.59 (7.90)	22.79 (6.81)	0.88

Table 28 (continued)

Means (Standard Deviations) and F Statistics for the Strong Vocational Interest Blank

Interest Area	High Level Performers	Average Level Performers	Low Level Performers	F
Musician Performer	27.70 (9.34)	30.09 (7.16)	31.21 (8.09)	0.94
Music Teacher	26.80 (9.49)	29.36 (13.02)	27.58 (7.68)	0.33
C.P.A. Owner	19.60 (11.24)	19.82 (9.52)	19.90 (7.54)	0.01
Accountant	23.70 (11.08)	25.59 (12.36)	21.11 (10.27)	0.80
Office Worker	30.60 (11.85)	32.86 (10.74)	29.21 (7.91)	0.66
Purchasing Agent	29.25 (10.07)	31.09 (9.07)	29.16 (5.98)	0.34
Banker	23.80 (10.11)	26.64 (12.98)	25.42 (12.68)	0.59
Pharmacist	27.00 (10.08)	25.55 (7.48)	27.79 (6.69)	0.40
Funeral Director	29.75 (10.12)	29.73 (6.75)	29.47 (8.210	0.01
Sales Manager	26.20 (11.05)	24.46 (7.63)	24.26 (8.20)	0.28
Real Estate Salesman	34.85 (11.36)	34.23 (5.61)	35.58 (8.80)	0.12
Life Insurance Salesman	27.95 (11.34)	27.14 (6.51)	28.79 (10.38)	0.15
Advertising Man	27.45 (9.14)	25.86 (8.52)	25.47 (8.84)	0.27
Lawyer	27.15 (10.76)	27.14 (7.14)	29.21 (7.73)	0.37
Author-Journalist	26.75 (9.34)	26.09 (8.34)	27.37 (6.18)	0.13
President-Manufacturing	18.10 (8.85)	17.09 (9.02)	14.74 (6.11)	0.87

Table 28 (continued)

Means (Standard Deviations) and F Statistics
for the Strong Vocational Interest Blank

Interest Area	High Level Performers	Average Level Performers	Low Level Performers	F
Credit Manager	34.35 (13.63)	35.14 (11.55)	32.00 (8.94)	0.40
Chamber of Commerce Exec.	36.95 (12.25)	37.41 (8.74)	34.36 (9.47)	0.51
Physical Therapist	41.35 (12.70)	42.14 (12.68)	44.47 (7.25)	0.40
Computer Programmer	24.85 (10.11)	26.41 (10.78)	25.84 (8.90)	0.13
Business Ed. Teacher	34.90 (13.24)	36.96 (9.69)	34.00 (8.14)	0.43
Community Rec. Admin.	37.50 (14.62)	39.36 (10.48)	37.37 (9.91)	0.19
Academic Achievement	38.85 (15.01)	39.82 (8.13)	40.79 (7.87)	0.16
Age Related	38.00 (12.04)	40.50 (7.98)	38.84 (9.96)	0.34
Diversity of Interests	54.75 (16.05)	61.50 (8.43)	57.37 (6.24)	2.00
Masculinity	58.85 (16.17)	60.36 (9.16)	60.90 (7.59)	0.17
Managerial Orientation	34.00 (11.40)	35.86 (9.31)	31.95 (9.36)	0.77
Introversion, Extroversion	35.90 (12.07)	39.78 (10.31)	41.21 (9.18)	1.31
Occupational Level	48.95 (12.55)	46.91 (6.13)	49.53 (5.93)	0.52
Specialization	37.40 (16.07)	39.00 (13.45)	37.37 (12.05)	0.15
Like Percentage	33.55 (16.07)	33.00 (13.45)	29.63 (12.05)	0.45
Indifferent Percentage	34.90 (18.91)	43.95 (16.26)	40.63 (18.23)	1.37

Table 29

Prediction Results
of the Strong Vocational Interest Blank

Actual Group Membership	Number of Cases	Predicted Group Membership		
		Group 1	Group 2	Group 3
Group 1 High Level	20	17 85.0%	2 10.0%	1 5.0%
Group 2 Average Level	22	0 0.0%	22 100.0%	0 0.0%
Group 3 Low Level	19	2 10.5%	0 0.0%	17 89.5%

Percent of Cases Correctly Classified: 91.80%.

are as follows:

F across all groups, F =2.61, p<.01

Group 1 and Group 2, F =3.34, p<.01

Group 1 and Group 3, F =1.81, p<.05

Group 2 and Group 3, F =2.92, p<.05.

These results indicate that when the following variables are taken in the order shown in Table 30 the three groups can be significantly discriminated from each other. The interaction of these variables serves well to separate these groups. It should be noted that a trend appears which

Table 30

Order of Variables Selected in SVIB Multiple Discriminant Analysis

STEP NUMBER	VARIABLE ENTERED*	REMOVED*
1	Art	
2	Religious Activities	
3	Physician	
4	Social Service	
5	Dentist	
6	Military Activities	
7	Music	
8	Law/Politics	
9	Business Management	
10	Librarian	
11	Writing	
12	Personnel Director	
13	Sales	
14		Business Management
15	Psychologist	
16	Mechanical	
17	Policeman	
18		Military Activities
19	YMCA Staff Member	
20		Physician
21	Public Speaking	
22	Teaching	
23	Biology	
24	Farmer	
25	Psychiatrist	
26	Engineer	
27		YMCA Staff Member

*"Entered" and "Removed" indicate which variables are being included in the grouping of variables that are used to discriminate performance levels.

suggests that those having higher interests in the professions (as represented by "dentist", "physician", and "biologist") tend to have lower performance ratings.

Cornell Index

The means, standard deviations, and F statistic for the one variable on this test are shown in Table 31. Although these results are not significant, they approach significance ($p < .10$). It is indicated that the high level performers tend to have fewer physical complaints than do the average and low level performers.

Hostility and Direction of Hostility Questionnaire (HDHQ)

The means, standard deviations, and F statistics for the two variables on this test that were considered ("direction of hostility" and "amount of hostility") are shown in Table 32. No significant differences occurred among the three groups on these test variables. It should be noted that the "direction of hostility" variable shows that all three groups direct hostility slightly outward (the "direction" score being greater than zero). It has been reported by the authors of this test (Caine, Foulds, and Hope, 1967) that normals tend to score in the extra-puntive direction as has occurred in these results.

Table 31

Group Means (Standard Deviations) and F Statistic for the Cornell Index

Cornell Variable	High Level Performers	Average Level Performers	Low Level Performers	F
Cornell Total	1.26 (1.28)	3.17 (3.05)	2.26 (2.58)	3.11

Table 32

Group Means (Standard Deviations) and F Statistics for the Hostility and Direction of Hostility Questionnaire

Hostility Variable	High Level Performers	Average Level Performers	Low Level Performers	F
Direction of Hostility	1.63 (3.40)	1.70 (4.45)	1.26 (3.00)	.08
Amount of Hostility	7.68 (3.71)	8.39 (4.14)	9.53 (5.64)	.80

Wechsler Bellevue Vocabulary

The means, standard deviations, and F statistics for the one variable from this test are shown in Table 33. No significant difference occurred across the three groups for this test variable. These group average scores fall within the average range according to the normative data supplied with this test.

California Psychological Inventory

The means, standard deviations and F statistics for the 18 variables of this test are shown in Table 34. These results indicate that only one of the 18 variables on this test significantly discriminated among the three groups. It is reported that this variable, "Capacity for Status", measures the "personal qualities and attributes which underlie and lead to status" (Gough, 1956, p. 10). The best policemen, as rated by their sergeants, score highest on this variable when compared with average and low performing policemen.

A multiple discriminant analysis was applied to the data for the 18 CPI variables. The analysis reached maximum separation among the three groups by using six of the 18 available variables. These variables were: "Capacity for status", "Intellectual efficiency", "Tolerance", "Achievement via independence", "Sense of well-being", "Social-

Table 33

Group Means (Standard Deviations) and
F Statistic for the Wechsler-Bellevue Vocabulary

Wechsler-Bellevue Vocabulary	High Level Performers	Average Level Performers	Low Level Performers	F
Raw Score	28.65 (3.86)	27.64 (3.76)	29.11 (4.04)	0.78

ization". The discriminant functions which were derived resulted in a correct classification of 62.30% of the three performance groups. The classifications as they occurred are shown in Table 35.

The F ratios for the discriminant functions developed are as follows:

F across all groups, F =2.29, p<.01
Group 1 and Group 2, F =2.22, p<.10
Group 1 and Group 3, F =2.97, p<.025
Group 2 and Group 3, F =1.80, p<.25.

The results indicate that the CPI only discriminates significantly (p<.05) between groups one and three. This test is best used as a more general screening device which will separate only the best performers from the poorest.

Table 34

Group Means (Standard Deviations) and F Statistics for the California Psychological Inventory

CPI Variable	High Level Performers	Average Level Performers	Low Level Performers	F
Dominance	57.37 (8.94)	54.57 (11.84)	53.00 (13.21)	.71
Capacity for Status	57.21 (6.78)	52.70 (7.63)	51.11 (8.01)	3.42*
Sociability	57.63 (7.33)	55.56 (8.03)	55.84 (7.78)	.42
Social Presence	60.74 (5.82)	56.61 (8.33)	57.95 (6.50)	1.81
Self-Acceptance	59.53 (7.19)	56.74 (8.10)	56.26 (8.51)	.10
Well-Being	57.90 (6.01)	58.52 (5.95)	55.21 (9.00)	.42
Responsibility	50.84 (7.40)	51.87 (7.49)	49.79 (7.21)	.42
Socialization	56.84 (6.04)	54.74 (8.72)	52.90 (8.43)	.27
Self-Control	57.47 (8.01)	57.30 (7.61)	55.74 (8.88)	.27
Tolerance	57.53 (6.19)	54.13 (7.09)	53.58 (7.63)	1.81
Good Impression	59.00 (11.45)	59.70 (10.34)	55.84 (10.80)	.72
Communality	56.05 (6.96)	55.43 (5.12)	57.78 (5.16)	.91
Achievement via Conformance	58.58 (6.45)	58.65 (6.21)	55.53 (9.81)	1.09
Achievement via Independence	58.21 (10.00)	54.48 (8.39)	53.95 (8.93)	.10
Intellectual Efficency	56.11 (7.30)	51.26 (9.50)	54.84 (8.56)	1.83
Psychological Mindedness	55.84 (6.90)	56.48 (7.08)	56.68 (7.46)	.07
Flexibility	48.00 (9.90)	49.09 (10.80)	49.11 (9.30)	.08
Femininity	46.84 (9.01)	49.74 (9.68)	46.84 (8.69)	.71

*$p < .05$

Table 35

Prediction Results of the
California Psychological Inventory

Actual Group Membership	Number of Cases	Predicted Group Membership Group 1	Group 2	Group 3
Group 1 High Level	20	12 63.2%	3 15.8%	8 21.1%
Group 2 Average Level	22	5 21.7%	14 60.9%	4 17.4%
Group 3 Low Level	19	2 10.5%	5 26.3%	12 63.2%

Percentage of Cases Classified Correctly: 63.30%.

The Situation Tests and the Three-Minute-Tests

Table 36 shows the means, standard deviations, and independent F statistics for each of the six tests. Only one of the six tests, "Directions", significantly discriminated among the three groups. The three-minute-test "World" was very close to significantly discriminating ($p < .10$) but because this test was only administered to the second recruit class the reduced sample size ($n = 28$) probably made the significance level more difficult to achieve.

Table 36

Group Means (Standard Deviations) and F statistics for the Situation Tests and the Three-Minute-Tests

Test Given	High Level Performers	Average Level Performers	Low Level Performers	F
Lawnchair	2.60 (0.70)	2.40 (0.39)	2.41 (0.87)	0.59
Directions	3.16 (0.47)	2.71 (0.52)	2.61 (0.90)	4.03*
Lifegaurd	2.69 (0.95)	2.46 (0.54)	2.42 (0.87)	0.84
3MT-People	2.68 (0.56)	2.55 (0.65)	2.36 (0.82)	0.97
Dog	3.14 (0.45)	2.89 (0.46)	2.96 (0.90)	0.62
3MT-World	3.37 (0.57)	3.12 (0.68)	2.45 (0.77)	3.23

*p<.025

The situation tests concerning areas of conflict ("Lawnchair", "Lifeguard", "Dog") did not successfully discriminate among the performance groups. The tasks which required verbalizing in the presence of little or no conflict ("Directions", "3MT-World") were the best tests. The candidate's ability to express abstract concepts, his level of maturity, and ability in verbal communication, are developmental areas more obviously observed on these tests.

The multiple discriminant analysis for the situation tests and the three-minute-tests achieved maximum separation among the groups with the use of only two of the six tests. The "Directions" situation test and the "World" three-minute-test were the two contributing variables used. The other tests added no additional unique discrimination power to the two discriminant functions derived. The predicted group membership as compared with the actual group membership (high level performers, average level performers, and low-level performers) is shown in Table 37.

The F ratios for the discriminant functions developed are as follows:

F across all groups, $F = 3.42$, $p < .01$
Group 1 and Group 2, $F = 2.82$, $p < .10$
Group 1 and Group 3, $F = 4.62$, $p < .01$
Group 2 and Group 3, $F = 2.96$, $p < .05$.

These results indicate that when the "Directions" and "World" tests are combined, groups one and three and groups two and three can be distinguished. This implies that future job performance at the middle and lower end of the job skill continuum can be predicted with these two tests. It is clear that group three can be discriminated from both of the other groups with this test.

General Analysis of The Best Test Variables

The groups of variables that were used in a multiple discriminant analysis for separate tests (Incomplete Sentence Blank, Draw-a-Person, Strong Vocational Interest

Table 37

Prediction Results of the
Situation Tests and the Three-Minute-Tests

Actual Group Membership	Number of Cases	Predicted Group Membership		
		Group 1	Group 2	Group 3
Group 1 High Level	20	9 45.0%	5 25.0%	6 30.0%
Group 2 Average Level	22	5 22.7%	9 40.9%	8 36.4%
Group 3 Low Level	19	7 36.8%	1 5.3%	11 57.9%

Percent of cases correctly classified: 47.54%

Blank, Califronia Psychological Inventory, Situation Tests and Three Minute Tests) plus individual test variables that had not been included in a previous multiple discriminate analysis (Wonderlic Total Score, Number Wrong on Wonderlic, Cornell Index Score, Total Hostility Score, Direction of Hostility Score, Wechsler Bellvue Vocabulary Score) were included in a large multiple discriminant analysis. The scores for the Thematic Apperception Test were not included because the direction of the differences appeared (insignificantly) to be in the wrong direction.

The prediction table for this analysis of the best variables in the test battery is shown in Table 38. Only one subject was misclassified when the larger analysis was applied to the data.

The F ratios for the discriminant functions developed are as follows:

F across all groups, F =4.12, $p<.01$
Group 1 and Group 2, F =4.43, $p<.01$
Group 1 and Group 3, F =3.67, $p<.01$
Group 2 and Group 3, F =4.31, $p<.01$.

Significant differences occurred between all pairs of groups and among the three groups.

The stepwise multiple discriminant analysis begins by selecting the single best discriminating variable according to the preselected criteria (sergeants' rankings in squad). Then a second variable is selected as the variable best able to improve the discrimination between the groups. The third and successive variables are selected according to their ability to furthur contribute to discrimination among the groups (See Nie, et al. Statistical Package for the Social Sciences, 1975, p. 436). The variables were selected in the following order as shown in Table 39. It can be noted that the best test to discriminate among the groups was the "Directions" situation test followed by "Ommissions" from the Incomplete Sentence Blank and "Arms" of the Draw-a-Person. These variables are very difficult to fake and

Table 38

Prediction Results for the Best Variables of the Test Battery

Actual Group Membership	Number of Cases	Predicted Group Membership		
		Group 1	Group 2	Group 3
Group 1 High Level	20	19 95.0%	0 0.0%	1 5.0%
Group 2 Average Level	22	0 0.0%	22 100.0%	0 0.0%
Group 3 Low Level	19	0 0.0%	0 0.0%	19 100.0%

Percent of Cases Correctly Classified: 98.36%.

probably for this reason are the best discriminators.

The variables shown in Table 40 are weighted so that the groups can be maximally separated. This separation occurred with the use of two discriminant functions. The plotting of the mean discriminant functions scores for the three groups is shown in Figure 1. These centroids (mean discriminant function scores) indicate that Group 1 and Group 3 are highly similar on Function 1 and quite different on Function 2. Group 2 differs from both of the other groups on both functions. If the concern in the selection process is primarily not selecting members from Group 3, then noting the individuals who scored highest on Function 2

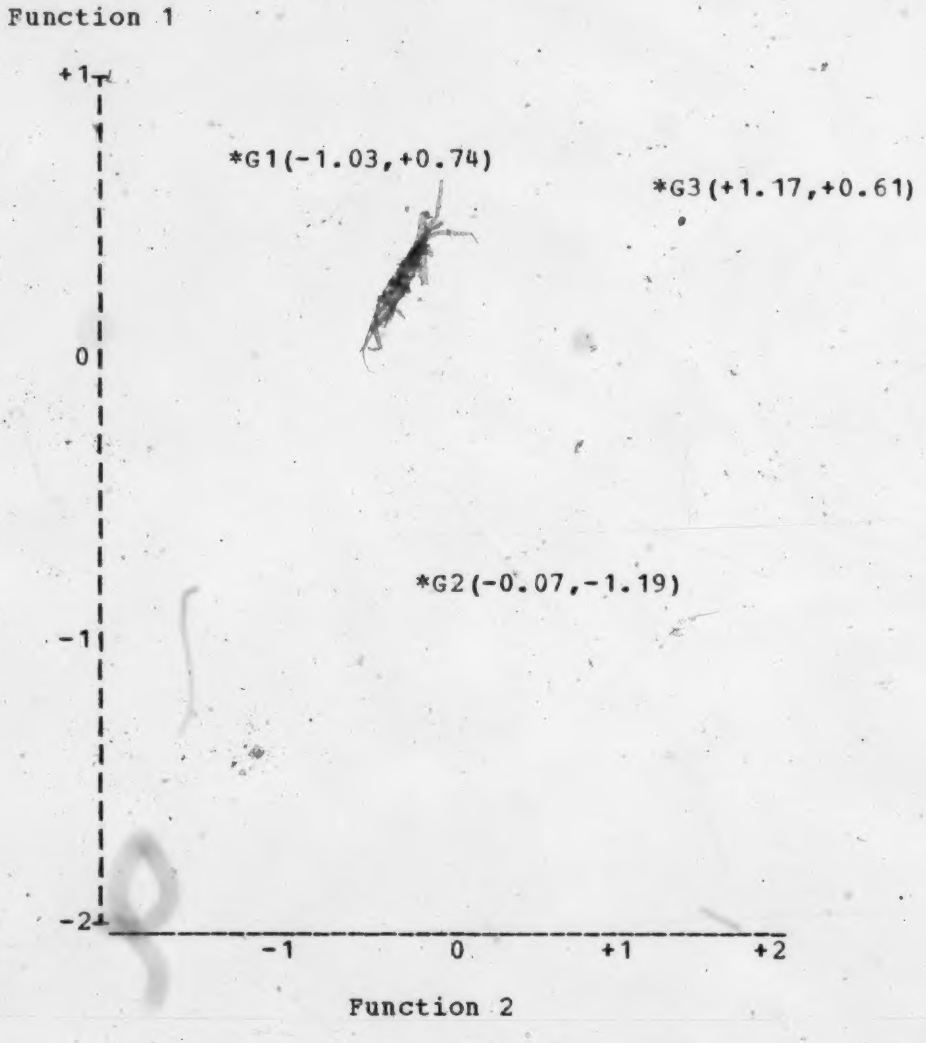
Table 39

Order of Variables Selected in Test Battery Discriminant Analysis

STEP NUMBER	VARIABLE	
	ENTERED	REMOVED
1	Directions, Situation Test	
2	Ommissions, Incomplee Sentence Blank	
3	Arms, Draw-a-Person	
4	Social Service, SVIB	
5	Sheet Position, Draw-a-Person	
6	Tolerance, CPI	
7	Sexual Identity, Draw-a-Person	
8	Waist, Draw-a-Person	
9	Psychiatrist, SVIB	
10	Line Quality, Draw-a-Person	
11	Achievement via Independence, CPI	
12	Well-Being, CPI	
13	Capacity for Status, CPI	
14	Socialization, CPI	
15	Mechanical, SVIB	
16	Physical Therapist, SVIB	
17		Social Service, SVIB
18	Psychologist, SVIB	
19	Librarian, SVIB	
20	Placement, Draw-a-Person	
21		World, 3MT
22	Sales, SVIB	
23	Liking Non-Human, Incomplete Sentence Blank	
24	Dentist, SVIB	
25	Personnel Director, SVIB	
26	Writer, SVIB	
27		Psychiatrist, SVIB
28		Sales, SVIB
29	Achievement by Independence, CPI	
30	Neck and Chin, Draw-a-Person	
31	Music, SVIB	
32	Facial Expression, Draw-a-Person	
33	Wonderlic Errors	
34		Shoulders, Draw-a-Person
35	Policeman, SVIB	
36	Identity, Draw-a-Person	
37	legs, Draw-a-Person	
38	Shoulder, Draw-a-Person	
39	Intellectual Efficiency, CPI	
40	World, 3MT	

Figure 1

Plotting of Group Centroids



would allow the best prediction power for this type of selection.

Defining what qualities are measured by the discriminant functions can be determined by observing the standardized discriminant weights (which are used as the multiplying coefficients to determine the contribution of each variable). These weights are shown in Table 40. High loadings on Function 1 show that an inverse relationship exists between Function 1 and "Librarian" (SVIB) and between Function 1 and "Arms" (Draw-a-Person). High scores on Function 1 occurred for both the high level performers and the low level performers. It is indicated that those who score high on Function 1 have low interests in the fields of "librarian", "dentist", and "psychologist" suggesting a lacking of inclination towards these types of professions. Function 2 has high positve weightings on the variables "physical therapist" (SVIB) and "dentist" (SVIB). Low scores on Function 2 occurred for the high level performers. It is indicated that high level performers have lesser interest in these areas then do the low level performers. This suggests a dedication to police work that is not interfered with by interests in other professions.

Table 40

Standardized Discriminant Function Coefficients for the two Discriminant Functions

VARIABLE	FUNCTION 1	FUNCTION 2
Directions, Situation Test	-0.15231	-0.17697
World, 3MT	0.07148	-0.17267
Sheet Placement, Draw-a-Person	0.43290	0.27331
Sex of Drawn Person, DAP	-0.45893	-0.07615
Line Quality, Draw-a-Person	-0.32751	-0.54505
Neck and Chin, Draw-a-Person	0.55769	-0.06902
Facial Expression, Draw-a-Person	0.23868	0.22560
Shoulders, Draw-a-Person	0.23868	0.22560
Arms, Draw-a-Person	-1.11066	-0.27686
Waist, Draw-a-Person	0.71308	-0.03054
Legs, Draw-a-Person	0.18942	0.28811
Liking Non Human, Incomplete Sentence Blank	0.45119	-0.19690
Response Omitted Incomplete Sentence Blank	0.14936	-0.51277
Mechanical Interest, SVIB	-0.73708	-0.62324
Music, SVIB	0.44694	0.01726
Writing, SVIB	1.02440	-0.19279
Dentist, SVIB	-0.93700	0.50827
Psychologist, SVIB	-0.93700	0.50827
Policeman, SVIB	0.52438	-0.03993
Personnel Director, SVIB	-0.91290	0.55943
Librarian, SVIB	-1.39356	0.35040
Physical Therapist, SVIB	-0.01412	0.81977
Number Wrong on Wonderlic	-0.17187	-0.13789
Capacity for Status, CPI	0.24101	-0.25565
Well Being, CPI	-0.03976	0.40340
Socialization, CPI	-0.10870	-0.29454
Tolerance, CPI	0.05923	-0.60844
Intellectual Efficiency, CPI	-0.45758	0.05985

CHAPTER IV

DISCUSSION

Intellectual Measures: Wonderlic Personnel Test and the Wechsler Bellevue Vocabulary.

It was hypothesized that the overall mean of the police recruits would be in the average or slightly above average range for these two tests. This hypothesis is not rejected. The mean scores for the police recruits ranged from 26.48 (high level performers) to 22.21 (low level performers). As predicted these scores fall within the average range, a score of 26-27 occurs at the 61.6 percentile, while a score of 22-23 occurs at the 39.2 percentile (see Wonderlic, 1961, p. 6).

The trend for the number correct, although not significant, suggests that high performers scored higher than average performers who scored higher than low performers. This trend has not been supported in the literature (National Crime Commission, 1928; Thweatt, 1972). Thweatt has reported that "the very bright men do not stick with the force" (p. 62). In this present study most, but not all dropouts, were placed by their job performance in the low performance group. This group had the lowest mean score on the intellectual measures.

The Wechsler Vocabulary Test score showed that no significant difference occurred among the groups. As predicted the mean raw scores (27-29) were within the above average range of intellectual functioning (average scale score of 12 on the Wechsler Bellevue, 80th percentile). These results suggest the following conclusions.

> (1) Policemen tend to be above the population average in intelligence.
>
> (2) Little relationship between intelligence test scores (of this above average intelligence group) and job performance exists.

The first conclusion substantiates a general finding that has been presented in the literature survey. For most police departments it is reported that policemen tend to be average to above average in intelligence (National Crime Commission, 1928; Kole 1962; Matarazzo, Allen, Saslow, and Wiens (1964); Neiderhoffer, 1967; Fenester and Locke, 1973; Matarazzo, Wiens, Matarazzo and Manaugh, 1973). The results of this study are congruent with this finding. It should be noted that these good intelligence scores could have been produced by the present selection method of the personnel department. A general knowledge questionnaire is given to each applicant and a 60% score is required before further consideration of the applicant is allowed. This screening method may be responsible for preventing those with below average intelligence from joining this police department.

Conclusion two also supports previous research. Little relationship between intellectual functioning of police recruits with above average mean intelligence scores and job performance has been demonstrated in the literature (Dubois and Watson, 1950; Mills, McDevitt, and Tonkin, 1966; Bass, Karstendiek, McCullaugh, and Pruitt, 1954). This result was also substantiated in this present study by the multiple discriminant analysis of the entire test battery. The variables "number right" on the Wonderlic and the "Wechsler Vocabulary Score" were not good discriminators of job performance levels. This was shown by the fact that these variables were not included in the final multiple discriminant analysis. It appears that once policemen are within the average to above average range of intellectual functioning it is other variables that determine the quality of their job performance.

The "number wrong" on the Wonderlic was the twenty fifth variable entered into the multiple discriminant analysis. This measure of carelessness and imprecision in intellectual tasks makes a useful but very limited contribution to discriminating among different levels of on-the-job performance.

The Interest Test:
Strong Vocational Interest Blank

It was hypothesized that the mean scores for all three groups on the "policeman" scale would be in the 'A' category (upper three quarters of the criterion group). This hypothesis cannot be entirely rejected. The "policeman" scale means fall within the 'B' category and are similar to one quarter of the original criterion group (Strong and Campbell, 1966). Cultural differences may explain why this present police group differed from the majority of the criterion group. It should be remembered that the criterion group was made up of American police officers. No significant differences among the three levels of performers for the "policeman" scale occurred. These results indicate that for this group of policemen the "policeman" scale was not a particularly good measure of job interest or of later job performance.

It was hypothesized that good policemen would score higher on mechanical and clerical measures. The results indicate that mechanical and clerical measures do not discriminate among performance levels of policemen. "Technical supervision" and "mechanical" interests showed no differences among the groups. In the literature survey, differences were shown in these areas on more performance oriented tasks rather than tasks such as the SVIB which is the paper and pencil type.

It was also hypothesized that high scores on professional scales ("dentist", "osteopath", "veterinarian", "biologist", "architect", "physician", "psychologist", "psychiatrist", "author journalist") and the "diversity of interest" scale would have a negative relationship with performance. This hypothesis is only partially rejected. It was shown that the scales "diversity of interests" (a measure of "breadth of interests"), "art", "dentist", "physician", and "biologist" were the best univariate discriminatiors although the differences were not significant. The low performers tended to have higher mean scores on "diversisty of interests" and the other professional scales. This finding in part supports the hypothesis that high interest scores in other professions relates negatively to police job performance measures. These findings substantiate those of Blum (1964) and Thweatt (1972). These authors reported that high scores in some professions had an inverse relationship with later police job performance measures. Thweatt (1972) described policemen in this category as "frustrated professionals". The results from this study may indicate that low level policemen may find police work frustrating and therefore report interests and aspirations in areas other than police work. These other interests may detract from their job performance.

Also hypothesized was the belief that high performing policemen would have a greater interest in social contacts. This hypothesis is not accepted. No significant results occurred in this area. For the scale "personnel director" the trend was in the direction predicted but the differences were far from significant.

It was hypothesized that the three police groups (the three groups being considered together) would score very high on the "masculinity-femininity" scale. This hypothesis is not rejected. This scale was developed by comparing the interests of men and women. "High scores on the Men's form indicate strong 'typically masculine' interests and low scores indicate 'typically feminine' interests" (Strong, 1966, p. 18). The original criterion group of policemen had a mean of 50 (standard deviation of 10). The present results show a mean of 60 (standard deviation of 11) indicating a unique interest area (which emphasizes masculinity) for this group as compared with men-in-general. For the purposes of prediction, this scale did not discriminate among performance levels.

The multiple discriminant analysis of the SVIB test data resulted in near perfect group prediction (91.80%). Twenty one variables were included in the final analysis, the best five variables being "art", "religious activities", "physician", "social service" and "dentist". The value of the multiple discriminant analysis can be seen in the selection of "art" as the best discriminating variable. This variable separated the average performers who scored higher than both the high level and low level performers.

The following general conclusions are suggested from the findings of the interest test.

(1) The "policeman" scale is not a particularly good measure of police interest or of later job performance.

(2) Policemen do not have higher interests in mechanical and clerical fields as measured by the SVIB.

(3) An inverse relationship between interest in other professions and job performance appears (not quite significantly) to exist.

(4) Policemen do not show a greater interest than men-in-general in social contacts.

(5) Policemen score very highly on items indicating "typically masculine" interests.

Personality Measures

Incomplete Sentence Blank

It was hypothesized that high level performers would score higher on the developmental scales ("Contribution to Society", "Feelings of Adequacy", "Liking for Others", "Others Like Subject") and lower on the non-developmental scales ("Harmful Generalization", "Destructive to Society", "Feelings of Inadequacy", "Disliking for Others', "Others

Dislike Subject"). No hypotheses were formulated for the scales "Harmless Generalization", "Self Serving", and "Response Omitted".

This hypothesis can not be rejected for two scales, "Feelings of Inadequacy" and "Destructive to Society". Although univariate F scores were insignificant (p<.30) the trend was in the expected direction and significant levels were approached. For the other developmental and non-developmental scales concerning the hypothesis in question no significant differences or trends were noted. These results suggest that high level performers have fewer feelings of inadequacy and fewer attitudes that are in opposition to societal values.

The "Response Omitted" variable, about which no hypothesis was formulated was the best variable on this test discriminating significantly (p<.05) among the three groups. The fact that high level performers omitted more sentence completions than did the other two groups allows for the following interpretations.

> (1) The high level performers are guarded and defensive about their personal life and are not willing to share it with strangers.
>
> (2) The "Response Omitted" variable is really a measure of carefulness and precision in answering questions. The high level performers demonstrated more concern in formulating their responses.
>
> (3) A time limit was applied to the test. The high level performers may have taken more time and considered more seriously each sentence stem and therefore did not omit questions but merely did

not have time to finish the test.

The first two interpretations appear more acceptable than the third. If the third interpretation was correct then the high level performers would have consistently omitted more sentence stems nearer the end of the test. This trend was not observed for the high level performers in the scoring procedure for the test. One study, reported in the literature survey, supports the first interpretation. Hooke and Krauss (1971) found that successful policemen (as measured by their promotion to sergeant) tended to be more guarded (higher scores on K, defensiveness; and Pa, paranoia; from the MMPI). Although this study was not supported by others in the literature comparable results appear to have occurred on this test.

The multiple discriminant analysis allowed for a prediction percentage of 47.54%. This indicates that the test as a whole was a relatively poor predictor of police performance.

Draw-a-Person

It was hypothesized that high level performers would score on this test to indicate greater maturity. This hypothesis is not rejected for only one variable, "Arms of Drawn Person". For this variable the high level and average level performers significantly ($p<.10$) were rated higher then the low level performers. This occurrence may have resulted if the appropriate drawing conception of "arms" is a reflection of an individuals self-concept. To score poorly on this variable the arms have to be drawn in an under or over-exaggerated manner. Both types of responses may reflect an inappropriate self-concept. If an individual's self-concept is reflected in this variable then this would have a definite relationship with job performance.

This test demonstrated its good potential to predict future job performance when the multiple discriminanat analysis statistical method was used. Only seven misclassifications occurred for the 61 subjects. This indicates that this test is difficult to "fake good". Further refinement of the test is needed so that distinct variables can be isolated that will discriminate performance groups.

Thematic Apperception Test

It was hypothesized for this test that the high performance policemen would score higher on the "Overall Level of Adjustment" scale for each TAT card. No significant differences occurred for any of the cards shown. The trend appeared to be in the opposite direction predicted although this trend was not significant. The scoring of the test was based upon the raters ability to determine the "credo" (fundamental belief) underlying each story told in response to the slide shown. It was presumed that the maturity level of this "credo" would be reflected in later job performance as a policeman. The following conclusions are suggested from these unpredicted results:

> (1) The "credo" is not representative of policemen's behavior and knowing the "credo" does not give information about future police behavior.
>
> (2) The "credo" is a reflection of the belief system of the recruits but this belief system was changed substantially by police experience.
>
> (3) The "credo" is not accurately presented in the stories written by the candidates. In their efforts to impress the examiners the "credos" presented were those the candidates thought were most appropriate and probably the most idealistic.

All three conclusions are difficult to accept. It was assumed that "credos" were reflected in the stories written. This assumption may be wrong. Conclusion two and three combined may have some validity. The "credos" during the recruitment period of the policeman may reflect an exaggerated idealistic concept of the way the candidate hopes to be. If this conclusion were accurate then no relationship between the candidate's exaggerated idealism and later job performance was shown. It appears that more research in the area of the "credo" is necessary before more substantial hypotheses can be considered.

California Psychological Inventory

It was hypothesized that the police recruits would fall within the average range on this test. In general all mean scores fall between standard scores of fifty five and sixty which is slightly above average and validates the hypothesis. Profiles that tend to be above average on every scale are subject to being considered as "faking good" profiles.

> The "fake good" profiles are above average on virtually all the scales. Their high point, as might be expected, is on the Gi (good impression) scale, a finding which sets the simulated profile apart from most profiles obtained under normal testing conditions. (Gough, 1969, p. 16)

The mean profiles for the three groups whose "fake good" conditions appear show the "Good Impression" score quite high (means = 59, 60, 56). It is possible that most of the questions on this test are being answered by the candidates so as to present their best image. However, this conclusion can only be made with caution. When interpreting mean profiles it must be remembered that differences among individuals are being ignored. The results of each profile would have to be considered individually to determine specific "fake good" profiles.

Only the "Capacity for Status" variable discriminated among the three groups. The high level performers scored higher on this variable than did the other two groups. This measure of "ambition" is a good predictor of future job performance. These results suggest the following conclusions.

> (1) The high average profile scores indicate the possibility of a "faking good" approach to the testing.
>
> (2) Ambition, the striving to succeed, is a quality measured by the CPI which discriminates between higher level and lower level performers.

The discriminant functions developed allowed for a prediction correct classification of 62.30% and significant discrimination between groups one and two and one and three resulted. Because of the "fake good" potential of this test it is probably only the most subtle items that discriminated among the groups. For a future police selection battery it is recommended that only the "Capacity for Status" scale be included from this test.

Cornell Index

This test was used primarily as a quick screening device to indicate personal and psychosomatic distrubances. It has been noted previously that policemen tend to have higher psychosomatic complaints (Los Angeles County Sheriff's Department, 1971). As predicted the high level performers tended to have a lower score than the other two groups, the difference approaching significance ($p<.10$).

This indicates that high level performers have fewer psychosomatic problems. In general, however, few problems of this nature were demonstrated by all of the groups. Because these men were thoroughly prescreened it is understandable that this highly select group would have few personal and psychosomatic problems. Better use of this test might be found earlier in the selection procedures.

Hostility and Direction of Hostility Questionnaire

It was hypothesized that these recruits would fall within the normal range on this test. This occurred, with all groups responding in a manner that demonstrated a normal amount of minimal outward hostile expression. It has been reported by the authors of the test that "normals" (as represented by a group of adult hospital employees and nonpsychiatric adult patients) also tend to score in the extrapunitive direction. As would be expected the trend of means, although insignificant, indicates that high level performers have a lower "amount of hostility" score than do the other groups. Because of the tests inability to discriminate among the performance groups a number of conclusions can be considered.

(1) Little difference among the performance groups exist on the amount of hostility or the form of expression that hostility is given.

(2) The "amount of hostility" that policemen have does not affect their job performance.

(3) The "amount of hostility" does affect police

performance but this hostility is developed differently according to other variables (interpersonal skills, assertiveness ability, maturity) after the recruit has been with the police force for a period of time.

(4) The Hostility and Direction of Hostility Questionnaire is a poor measure of hostility.

Conclusions one and two are rejected. Authoritarianism is known to exist in police behavior and reportedly has an effect on job performance (see Westley, 1970, p. ii, p. 239; Toch and Schulte, 1961; Biles and McCoy, 1973; Teasley and Wright, 1973; Neiderhoffer, 1967; Barocas, 1973; Crawford, 1973; Skolnick, 1968, p. 9). Conclusion three may be correct. Hostility may have developed within the police working environment (after the testing) as a function partially of previous skills in handling hostility plus other variables. (These other important variables could include the modelling effect of the interaction style of the supervising sergeant, the police partners attitude toward citizens, the area in which the police work is located, etc.) This conclusion would account for the slight differences (although insignificant) that occur among the groups.

Conclusion four may also be partially acceptable. Any paper and pencil test in a selection procedure is "fakeable". The validity of the test may have been lessened because of the nature of the testing situation. It would be enlightening if an accurate measure of hostility could be given biannually to this police group to monitor amount and direction of hostility scores.

Situation Tests and Three-Minute-Tests

These tests were the best tests given to the police group, if one considers that very good discrimination was achieved on two of the six tests (Directions, 3MT-World). Both tests indicate maturity level and verbalizing skill in a difficult situation. The "Directions" test is quite simple. An actor asks a police applicant for directions to city hall (from the legislative grounds) in the presence of videotaping equipment and two other experimenters. The ability to verbalize and explain the correct route in a courteous and friendly manner appears to be a skill that most should be able to demonstrate. Yet it discriminates extremely well among the performance groups. This skill may represent a policeman's ability, to fulfill the public-servant role, to be a friend to a stranger in need of help. The skill in this role may be one prerequisite to being a good policeman.

The other good test, "3MT-World", required the police candidate to speak for three minutes in answering the question "What's wrong with the world?" Communicative ability and attitudes concerning many areas (governments, world problems, people, violence, city life, etc.) were presented. The "Level of Development" of the candidate as demonstrated on this test may reflect quite accurately his "mature perspective".

The situation tests dealing with the handling of hostile individuals were expected to show that those with skills in these areas would later demonstrate proficiency as policemen. The mean scores were generally in the right direction but the differences were not significant. It may be that motivation to succeed ("Capacity for Status" as measured by the CPI) is a much more important variable than ability to handle hostility. A highly motivated police recruit may overcome his poor skills in handling hostility as demonstrated during these selection procedures.

The results from the situation tests and the three-minute-tests suggest the following conclusions.

> (1) The ability to be friendly and helpful to a stranger, to be able to incorporate the public servat role, is a quality that high level policemen have more often than low level policemen.
>
> (2) Communication skills as represented by the ability to speak at length on a topic with no preparation, and at the same time demonstrate a mature value system, is a basic prerequisite of high level policemen. It may represent an individual's maturity and perspective.
>
> (3) Differences in abilities to handle hostile situations were not demonstrated to be related to later performance on-the-job. It appears that interpersonal skills in these areas are primarily developed in training or on the job. Situational problems highly comparable to job situations might be better used as a teaching method during training rather than as a selection technique.

Multiple Discriminant Analysis of Test Battery

When the entire test battery was entered into the multiple discriminant analysis, almost perfect predicted group membership occurred when the subjects were classified in terms of two discriminant functions. The contribution of the variables that were used in these functions may be determined by observing the standardized discriminant weights. These weights were used to calculate the relative contribution of each variable to the discriminant function. In understanding what is being measured by a discriminant function it is wise to consider the high positive and high negative weights for that function and from this information attempt to derive the essence of what is being measured.

Function 1. The high positive and high negative weights for Function 1 are shown in Table 41. The high positive weights indicate that individuals who score higher on these variables score highly on Function 1. (The high level performers and low level performers score highly on Function 1.) A high interest in writing and liking things non-human occur for Function 1 and suggest that high scorers tend to have intellectual interests and prefer being at a distance from people. High scores on "neck and chin" and "waist" variables from the Draw-a-Person are indicative of

Table 41

High Positive and High Negative
Weights for Function 1

High Positive Weights		High Negative Weights	
Writing	+1.024	Librarian	-1.390
Waist	+0.713	Arms	-1.390
Neck-Chin	+0.557	Dentist	-0.937
Like: Non Human	+0.451	Psychologist	-0.937

good self-concepts. Interests and attitudes of this nature may indicate that high scores in these areas reflect an intellectual, conservative, distancing approach to solving problems.

The high negative weights shown on Function 1 also indicate a desiring of distance from individuals. Those who want to help others and be involved with others in a professional way may choose careers such as "librarian", "dentist", and "psychologist". Individuals scoring highly on Function 1 tend to score low in these interest areas. A low score on the "arms" variable from the Draw-a-Person would also be achieved by those that score highly on Function 1. Low scores on the "arms" variable often result because the arms have been categorized (by the three raters) as "exaggerated, muscles shown" or "thin, weak, lopsided". The over or under-emphasis on arm strength may reflect feelings of inadequacy expecially in skills involving the use of the arms. Because of these self-concepts of physical inadequacy careers distant from others are sought.

Function 1 appears to be setting apart individuals who have intellectual interests ("writing", SVIB) and also desire good presentation ("neck and chin", "waist", DAP). Also indicated is the seeking of some distance from others ("liking: nonhuman", and lower interests in helping professions, "dentists", "psychologist", "librarian").

The high performing police group and low perfoming police group scored high on Function 1. This means that the "distancing" approach to policework is effective for the high level performers and is not for the low level performers. Many would suggest that a distant, disinterested approach to police work would be an effective working approach to the role of policemen. Consider Stirling's (1972) research where it is pointed out that policemen perceive fellow police officers and the courts as having role expectations requiring policemen to be objective (impersonal) and reserved in their approach. This orientation within police work would allow for a more judicious attitude in dealing with citizens. However, with this approach, compassion is often denied and the possibility of cynicism occurring is present.

The degree to which an individual remains distant and objective (Function 1) and yet not cynical and uninvolved may be reflected in Function 2. It is Function 2 that discriminates among the three groups the most clearly.

Function 2. The high postive weights and high negative weights for Function 2 are shown in Table 42. In explaining Function 2 it must be remembered that the high level performers have a low average score (-1.03), average level performers have a middle range average score (-0.07), and low level performers have a high average score (+1.17) on this function. The high positive weights on the scales of "physical therapist", "personnel director", "psychologist", and "dentist" may indicate aspirations for higher status professions. Intersts in these areas are more indicative of the low level performers than the high level performers. This occurrence may indicate that low level performers desire positions in other professions and are not highly movitivated to remain with police work.

One high negative weight that occurred for this function was "mechanical" interest (SVIB). This implies that the high level performers have more interest in this area and may be seeking hobbies or frustration outlets in non-people activities. The frustrations referred to may have resulted from the high "tolerance" score (CPI) which is also negatively weighted on this function. This indicates that the high level performers would tend to score highly on this scale from the CPI. This may be necessary for all high level policemen. The ability to tolerate frustration and

Table 42

High Positive and High Negative
Weights for Function 2

High Positive Weights		High Negative Weights	
Physical Therapist	+0.820	Mechanical	-0.623
Personnel Director	+0.560	Tolerance	-0.623
Psychologist	+0.508	Response Omitted	-0.513
Dentist	+0.508		

anxiety are extremely important in police work. This ability may also be reflected in the high negative weighting for the "Response Omitted" variable (Incomplete Sentence Blank). High level performers have had more responses omitted than the other groups. This may indicate an ability to cope with the possible negative consequences of such action (suspected negative reinforcement from the examiner).

Function 2 appears to be a measure of motivation for police work (as indicated by relatively lower interests in comparison with poor performers in other professions) and also a measure of tolerance in interpersonal and conflict situations. It is possible that a relationship between "motivation" and "tolerance" exists. If police are to tolerate frustration daily a high degree of dedication would be required.

The following conclusions are suggested from the multiple discriminant analysis of the entire test battery.

> (1) Function 1 appears to measure the relating style of an individual, the high level (and low level) policemen are more desirous of a distant objective relationship.
>
> (2) Function 2 appears to measure both motivation for police work and tolerance for frustration.
>
> (3) If a police candidate demonstrates a personality type that shows job dedication and an ability to tolerate frustration coupled with a professional approach to police work that maintains a certain distance between the policeman and the public then it appear that this candidate would be a good prospect for police work.

Concluding Remarks

From the statistical procedure of the multiple discriminant analysis areas that are defined by combinations of variables suggest the next steps in research projects dealing with problems of police selection. It appears that interests in intellectual activities, high motivation for police work, and tolerance for stress are qualities that are predicitive of good police job performance (as assessed by sergeants). Future research projects could attempt to more clearly define elements in these areas so that better predictions can be made when assessing policemen candidates.

A concluding remark on the situation tests is warranted at this time. In the literature situation tests were characteristically poorly standardized. Within this research project that problem was solved with the use of an actor working from a script. It was shown that situation tests could be used in a reliable manner. There was also evidence that these tests appeared to be accurately measuring a police candidates' ability to react intelligently and maturely under stress. Some of these measures were predictive of future job performance. Many of the police candidates remarked after the testing that they believed that the situation testing, although unsettling, was the fairest assessment of their abilities when all of the test battery was considered. More research in the area of the situational test would be of value for police personnel officers.

Future research projects in the police department for which this study was completed should also be considered. A revised test battery made up of the best predictor variables found in this study should be examined in a follow-up study to validate the present findings. This approach would also help validate the findings of this research.

Very few studies in the police literature have related test measures to a criterion. This research has demonstrated that relationships exist between psychological test scores of police applicants and performance on the job two years later. Seldom has it been shown that police applicants' test scores can be related to job performance measures. The value of this research could be further enhanced if the subjects in this study were followed throughout their entire police career.

This study has been conducted with the cooperation of the City of Edmonton Police Department, Edmonton, Alberta. This Canadian societal context is different in some aspects from many of the large American cities upon which articles reviewed from the literature were based. The unique aspects of this environment have an influence upon the job performance of policemen. This factor should be considered when generalizing the findings to areas that are vastly different from Edmonton, Alberta.

It has been demonstrated that certain psychological test measures are related to later job performance. It must not be forgotten that the job performance measure was derived from the rank-ordering of police recruits by their respective sergeants. The value of this study is dependent upon the accuracy of this measure.

Merits of this Study

This study had made the following contributions.

(1) A discriminant function has been developed so that a new recruit's test scores can be weighted and a prediction for future job performance can be made.

(2) Situation tests were developed so that police candidates reactions under stress could be measured in a standardized and reliable fashion. Two of these tests were significantly related to later job performance.

(3) This is one of the few studies in police research that is predictive in nature. A measure of job performance two years after the original test battery was given served as the criterion upon which the value of the tests were measure.

(4) This is a study centered in a Canadian city. Very few police studies have dealt with Canadian policemen.

Within this study psychological test results of sixty-one subjects (police recruits), gathered during the first week on-the-job, were compared with job performance measures two years later. These performance measures were derived from squad sergeants' assessments of the job performance of the subjects. Certain relationships between the psychological variables and the job performance scores were discovered. If these relationships can be validated in furthur studies, police job performance will be more easily predicted, and these predictions may be made at the time of hiring.

REFERENCES

Alberta Bureau of Statistics and the Edmonton Police Commission. Residents' attitudes towards the police. City of Edmonton Printing Department, August, 1976.

Alex, Nicholas Black in blue. New York: Appleton-Century-Crofts, 1969.

Anastasi, A. Psychological testing. (Third edition), New York: MacMillian, 1968.

Ansbacher, H. L. Murray's and Simoneit's (German military) methods of personality study. Journal of Abnormal and Social Psychology, 1941, 36, 589-592.

Ansbacher, H. L. Lasting and passing aspects of German military psychology. Sociometry, 1949, 12, 301-312.

Ansbacher, H. L., & Nichols, K. R. Selecting the Nazi officer. Infantry Journal, 1941, 49(11), 44-48.

Azen, S. P., Snibbe, H. M., & Montgomery, H. R. A longitudinal predictive study of success and performance of law enforcement attitudes. Journal of Applied Psychology, 1973, 57(2), 190-192.

Bad guys cop out---Now they are nice. Edmonton Journal, December 14, 1976.

Baehr, M. E., Furcon, J. E., & Froemel, E. C. Psychological assessment of patrolmen qualifications in relation to field performance. Washington, D. C.: U. S. Government Printing Office, 1968.

Bain R. The policeman on the beat. The Scientic Monthly, 1939, 48, 452.

Barocas, H. Urban police: Crisis mediators or crisis creators? American Journal of Orthopsychiatry, 1973, 43(4), 632-639.

Bass, R. M., Karstendiek, B., McCullaugh, G., & Pruitt, R. Validity information exchange. Personnel Psychology, 1954, 7, 159-160.

Beat police to increase. Edmonton Journal, January 4, 1977.

Berman, I. R., Darley, J. G., & Patterson, D. G. Vocational interest scales. University of Minnesota Bulletin, Employment Stabilization Research Institute, 1934, 3, 215-245.

Bersoff, D. N., & Greiger, R. M. An interview model for the psychosituational assessment of children's behavior. American Journal of Orthopsychiatry, 1971, 41, 483-493.

Biles D., & McCoy B. Police attitudes to deviance in Victoria. Australian and New Zealand Journal of Sociology. 1973, 9(2), 67-70.

Bittner E. The functions of the police in modern society. Maryland: National Institute of Mental Health, 1960.

Black, D. J. Police encounters and social organization: An observational study. Unpublished Ph.D. disseration. University of California, Berkeley, 1968.

Blum, M. L. & Naylor, J. C. Industrial psychology: Its theoretical and social foundations. New York: Harper and Row, 1956.

Blum R. H. Police selection. Springfield: Thomas, 1964.

Brenner, A. R. & Levin, J. M. Off duty policemen and bystander "apathy". Journal of Police Science and Administration, 1973, 1(1), 61-64.

Bruch, M. A. Psychological screening inventory as a predictor of college student adjustment. Journal of Consulting and Clinical Psychology, 1977, 45(2), 237-244.

Burnham, D. Police violence: A changing pattern. The New York Times, 7 July 1968.

Caine, T. M., Foulds, G. A., & Hope, K. Manual of Hostility and Direction of Hostility Questionnaire. London: University cf London Press, 1967.

Campbell, D. P. Strong Vocational Interest Blanks Manual (1969 Supplement). Stanford, California: Stanford University Press, 1971.

Campbell, J. P., Dunnette, M. D., Arrey, R. D., Hellervik, L. V. Journal of Applied Psychology, 1973, 57(1), 15-22.

Carte, G. E. Changes in public attitudes toward the police: A comparison of 1938 and 1971 surveys. Journal of Police Science and Administration, 1973, 1(2), 182-200.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. K. Sixteen Personality Factor Questionnaire (16PF). Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Chenoweth, J. H. Situational tests: A new attempt at assessing police candidates. Journal of Criminal Law and Police Science, 1961, 52, 232-238.

Collarelli, N. J. & Siegel, S. M. A method of police personnel selection. Journal of Criminal Law, Criminology and Police Science, 1964, 55(2), 287-289.

Crawford, T. J. Police over-perception of ghetto hostility. _Journal of Police Science and Administration_, 1973, _1_ (2), 168-174.

Cronbach, L. J. _Essentials of psychological testing_. New York: Harper, 1949.

Cronbach, L. J. Assessment of individual differences. _Annual Review of Psychology_, 1956, _7_, 173-196.

Cronbach, L. J. _Essentials of psychological testing_. New York: Harper and Row, 1970.

Cross, A. C. & Hammond, K. R. Social differences between "successful" and "unsuccessful" State highway patrolmen. _Public Personnel Review_, 1951, July, 159-161.

Cumming, E., Cumming, I., Edell, L. Policeman as philosopher, guide, and friend. _Social Problems_, 1965, _12_, 276-286.

Dahlstrom, W. G., & Welsh, G. S. _An MMPI handbook, a guide to use in clinical practise and research_. University of Minnesota Press: Minneapolis, 1962.

Darley, J. M. & Batson, C. D. Test of Samaritan Parable: Who helps the helpless? New York Times, April 11, 1971.

Devine, Frank S. Survey report number two for the Nebraska state patrol. Lincoln, Nebraska: mimeographed, not dated. See Stirling, J. W. Changes in role concepts of police officers. Gaithersburg, Maryland: International Association of Chiefs of Police, 1972.

Diamond, M. J. & Lobitz, W. C. When familarity breeds respect: The effects of an experimental depolarization program on police and student attitudes towards each other. Journal of Social Issues, 1973, 29(4), 95-109.

Diehl, H. S., & Patterson, D. G. A personnel study of Duluth policemen. Employment Stabilization Research Institute Bulletin, 1933, 2(2), 24-54.

Diehl, H. S., Patterson, D. G., Dorvak, G. J., & Longstaff, H. P. A personality study of Duluth policemen. Bulletin of the Employment Stabilization Research Institute, (University of Minnesota), 1933, II, 2, 35-54.

Dillman, E. G. Analyzing police recruitment and retention problems. Police, 1964, 8(5), 22-26.

Dubois, P. H., & Watson, R. I. The selection of patrolmen. Journal of Applied Psychology, 1950, 34, 90-95.

Dubois, P. H., & Watson, R. I. Validity information exchange. Personnel Psychology, 1954, 7, 414-416.

Earle, H. Stress versus non-stress training. Springfield: Thomas, 1973.

Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16(4), 107-424.

Eldefonson Edward (Ed.) Readings in criminal justice. New York: Collier-Macmillan, 1973.

Fenster, C. A., & Locke, G. The "dumb cop": Myth or reality? An examination of police intelligence. Journal of Personality Assessment, 1973, 37(3), 276-281.

Ferdinand, T. N. From a service to a legalistic style police department: A case study. *Journal of Police Science and Administration*, 1976, *4*(3), 302-305.

Ferguson, N. M. *Application documents and perfromance ratings of community police*. Unpublished master's thesis, University of Alberta, 1974.

Fernald, G. G. An achievement capacity test. *Journal of Educational Psychology*, 1912, *3*, 331-336.

Finney Robert D. *The bluecoat syndrome: An appraisal of police officers' perception of themselves*. Social Research Corporation of Utah: mimeographed, not dated. See Stirling, J. W. *Changes in role concepts of police officers*. Gaithersburg, Maryland: International Association of Chiefs of Police, 1972.

Fiske, D. Why do we use situation performance tests? *Personnel Psychology*, 1954, *7*, 464-469.

Fitts, P. M. German applied psychology during World War II. *Occupational Psychology London*, 1946, *21*, 151-161.

Flannagan, J. C. Some considerations in the development of situation tests. Personnel Psychology, 1954, 7, 461-464.

Fraser, A. W. A study of the standards and methods of selection used by the R.C.M.P. Unpublished Master's Thesis, Universisty of Alberta, 1949.

Frost, T. M. Selection methods for police recruits. Journal of Criminal Law and Criminology, 1955, 46, 135-145.

Furcon, J. Occassional paper 35, some questions and answers about police officer selection testing. University of Chicago: Industrial Relations Center, 1972.

Galton, F. Measurement of character. The Fortnightly Review, 1884, 36, 179-185.

Garmire, B. I. The police role in an urban society. In R. F. Steadman (Ed.), The police and the community. Baltimore: John Hopkins University Press, 1972.

Germann, A. C. Police personnel management. Springfield, Illinois: Thomas, 1958.

Ghiselli, E. E. & Brown, C. W. Personnel and industrial psychology. Toronto: McGraw-Hill, 1948.

Gottesman, J. Personality patterns of urban police applicants as measured by the MMPI. Hoboken, New Jersey: Laboratory of Psychological Studies, 1969.

Gough, H. G. The Adjective Check List. Palo Alto, California: Consulting Psychological Press, 1952.

Gough, H. G., McKee, M. G., & Yandell, R. J. Adjective Check List analyses of a number of selected psychometric and assessment variables. Technical memorandum, OERL-TM-55-10. Maxwell Air Force Base, Alabama: Officer Education Research Laboratory, 1955.

Gough, H. G. California Psychological Inventory. Palo Alto, California: Consulting Psychological Press, 1956.

Gough, H. G. California Psychological Inventory Manual, Palo Alto, California: Consulting Psychologists Press, 1969.

Hanley, D. M. Forced choice performance evaluation in a municipal police department. The Police Chief, 1975, January, 34-35.

Hankey, R. Personality correlates in a role of authority: The police. Dissertation Abstracts, 1968, 29(3-A), 952.

Hansen, D. A. An analysis of police concepts and programs. Springfield: Thomas, 1972.

Hartshorne, H., & May, M. A. Studies in deceit. New York: MacMillan Company, 1928.

He gets to know people walking the beat. Edmonton Journal, December 14, 1976.

Heiman, M. F. The police suicide. Journal of Police Science and Administration, 1975, 3(3), 267-273.

Hogan, R. Personality characteristics of highly rated policemen. Personnel Psychology, 1971, 24, 679-686.

Holmes, B. Selection of patrolmen. Journal of Criminal Law, Criminology and Police Science, 1942, 32(5), 557-558.

Holmes, J. A. Predicting success for security policemen. Personnel Psychology, 1951, 4, 397-403.

Hooke, J. F., & Krauss, H. H. Personality characteristics of successful police sergeant candidates. Journal of Criminal Law, Criminology and Police Science, 1971, 62 (1), 104-106.

Humm, D., & Humm, M. Humm-Wadsworth Temperament Scale appraisals compared with criteria of job success in the Los Angeles police department. Journal of Psychology, 1950, 30, 63-75.

Jenkins, W. O. A review of leadership studies with particular reference to military problems. Psychological Bulletin, 1947, 44, 54-79.

Johansson, C. B., & Flint, R. T. Vocational preferences of policemen. Vocational Guidance Quarterly, 1973, 22(1), 40-42.

Kates, S. Rorschach responses, Strong Blank scales, and job satisfaction among policemen. Journal of Applied Psychology, 1950, 34, 249-254.

Kelly, E. L. The place of situation tests in evaluating clinical psychologists. Personnel Psychology, 1954, 7, 484-492.

Kelly, R., & Farber, M. Identifying responsive inner-city policemen. Journal of Applied Psychology, 1974, 59 (3), 259-263.

King, P., Norrell, G., & Erlandson, F. L. The prediction of academic success in a police administration curriculum. Educational and Psychological Measurement, 1959, 19, 649-651.

Kole, D. M. A study of intellectual and personality characteristics of medical students. Unpublished master's thesis, University of Oregon Medical School, 1962.

Lane, R. Policing the city: Boston, 1822-1885. Cambridge: Harvard University Press, 1967.

Landy, R. J., Farr, J. L., Saal, F. E., Freytag, W. R. Behaviorally anchored scales for rating the performance of police officers. Journal of Applied Psychology, 1976, 61(6), 750-758.

Lawlis, G. F., & Lu, E. Judgment of counseling process: Reliability, agreement and error. Psychological Bulletin, 1972, 78(11), 17-20.

Leiren, B. O., Kiker, V. L., Phelan, J. G. Validating the selection of deputy marshals. Proceedings of the Annual Convention of the American Psychological Association. 1971, 6(1), 131-132.

Levy, R. Predicting police failures. Journal of Criminology and Police Science, 1967, 58(2), 265-267.

Lewis, R. W. Toward an understanding of police anomie. Journal of Police Science and Administration, 1973, 1 (4), 484-490.

Los Angeles County Sheriff's and the Los Angeles County Department of Personnel, Occupational Health Service. Physiological Fitness Standards, National Institute of Law Enforcement and Criminal justice, Annual Report, p. 2, June, 1971.

Machover, L. Machover Draw-A-Person Test, Springfield, Illinois: Charles C. Thomas, 1949.

Mann, P. A. Psychological consulation with a police department. Springfield: Thomas, 1973.

Marsh, S. H. Validating the selection of deputy sheriffs. Public Personnel Review, 1962, 23(1), 41-44.

Martin, E. M. Aptitude test for policemen. Journal of Crime and Criminology, 1923-1924, 14, 376-404.

Matarazzo, J. D., Allen, B. V., Saslow, G., & Wiens, A. N. Characteristics of successful policemen and firemen applicants. Journal of Applied Psychology, 1964, 48 (2), 123-133.

Matarazzo, R. G., Weins, A. N., Matarazzo, J. D., & Manaugh, T. S. Test-retest reliability of the WAIS in a normal population. Journal of Clinical Psychology, 1973, 29, 194-197.

May, M. A., & Hartshorne, H. Objective methods of measuring character. Pedagogical Seminary, 1925, 32, 45-67.

May, M. A., & Hartshorne, H. First steps toward a scale for measuring attitudes. _Journal of Educational Psychology_, 1926, _XVII_(3), 145-162.

McAllister, J. A study of the prediction and measurement of police performance. _Police_, 1970, _14_(14), 58-64.

McClelland, D. Testing for competence rather than intelligence. _American Psychologist_, 1973, _1_(1), 1-14.

McConnel, W. A. _Relationship of personal history to success as a police patrolman_. Unpublished Master's Thesis, Colarado State University, 1967.

McManus, G. P., Griffin, J. I., Wetteroth, W. J., Boland, M., & Hines, P. T., _Police training and performance study_. Washington, D. C.: United States Government Printing Office, 1970, PR-70-4, Grant No. 339, 75-80.

Merrill, M. A. Intelligence of policemen. _Journal of Personnel Research_, 1926-1927, _5_, 511-515.

Mills, R. B., McDevitt, R. J., & Tonkin, S. Situational tests in metropolitan police recruit selection. _Proceedings of the 73rd Annual Convention of the American Psychological Association_, 1965, 243-244.

Mills, R. B., McDevitt, R. J., & Tonkin, S. Situational tests in metropolitan police recruit selection. *Journal of Criminal Law, Criminology and Police Science*, 1966, *57*(1), 99-106.

Moore, H. T., & Gilliland, A. R. The measurement of aggressiveness. *Journal of Applied Psychology*, 1921, *5*, 97-118.

Moreno, J. L. The situation test in American-British military psychology vs. German military psychology. *Sociometry*, 1949, *12*, 344-353.

Morman, R. R., Hankey, R. O., Heywood, H. L., & Liddle, L. R. Predicting state traffic officer cadet academic performance from theoretical TAV selection system scores. *Police*, 1966(a), *10*(3), 54-58.

Morman, R. R., Hankey, R. O., Heywood, H. L., Rodgers, T., & Goldwhite, M. Multiple prediction of municipal police officers' ratings and rankings using the theoretical TAV selection system and certain non-test data. *Police*, 1967, *11*(3), 19-22.

Morman, R. R., Hankey, R. O., Kennedy, P., Heywood, H. L. Predicting state traffic officer performance with TAV Selection System Theoretical Scoring Keys. _Police_, 1965 (May-June), _5_, 70-73.

Morman, R. R., Hankey, R. O., Kennedy, P. K., and Jones, E. M. Academy achievement of state traffic officer cadets related to TAV selection systems plus other variables. _Police_, 1966(b), _10_(6), 30-34.

Motowidlo, S. J. and Borman, W. C. Behaviorally anchored scales for measuring morale in military units. _Journal of Applied Psychology_, 1977, _62_(2), 177-183.

Muir, W.K. _Police: Street corner politicians_. Chicago: University of Chicago Press, 1977.

Mullineaux, J. E. An evaluation of the predictors used to select patrolmen. _Public Personnel Review_, 1955, _16_, 84-86.

Murphy, J. P. Current practices in the use of psychological testing by police agencies. _Journal of Criminal Law, Criminology and Police Science_, 1972, _63_(4), 570-576.

Murray, H. A. Thematic Apperception Test, Cambridge, Massachusetts: Harvard University Press, 1944.

Murray, H. A., & Stein, M. Note on the selection of combat officers. Psychosomatic Medicine, 1943, 5, 386-391.

Murray, H. A., & MacKinnon, D. W. Assessment of OSS personnel. Journal of Consulting Psychology, 1946, 10, 76-80.

Murray, H. A. and others. Assessment of men: Selection of personnel for the Office of Strategic Services. New York: Rhinehart and Company, 1948.

Narrol, H. G., & Levitt, E. C. Formal assessment procedures in police selection. Psychological Reports, 1963, 12 (3), 691-694.

National Crime Commission's Report on the abstract intelligence of police officers. Public Personnel Studies, 1928, 6, 74-80.

Neiderhoffer, A. Behind the shield. New York: Doubleday, 1967.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., Bent, H. Statistical Package for the Social Sciences, 1975.

Nowicki, S. A study of the personality characteristics of successful policemen. Police, 1966, 10(3), 39-41.

O'Connor Wiggly Block, Chicago: C. H. Stoelting Company, 1951.

Oglesby, T. W. Use of emotional screening in the selection of police applicants. Public Personnel Review, 1957, 18, 228-231.

Ostrander, J. M. One hundred and fifty policemen, some observations upon 150 probationary policemen of the Detroit police department. Mental Hygiene, 1925, 9, 60-73.

Otis, A. R. Otis Quick Scoring Mental Abilities Test. New York: World Book Company, 1939.

Parratt, S. D. A scale to measure effectiveness of police functioning. Journal of Criminal Law and Criminology, 1938, 28, 739-760.

The Police Function in our Changing Society. Proceedings of Conference 'A', 'The role of the policeman', Lake Couchiching, Ontario, Canada: Consultation Center, Ministry of the Solicitor General, Ottawa, Ontario.

Police power: Is there too much? Canada and the World. 1974, December, 4-5.

Preiss, J., & Ehrlich, H. An examination of role theory. Lincoln, Nebraska: The University of Nebraska Press, 1966.

Police to promote senior power. Edmonton Journal, January 23, 1977.

Purviance, C. W. An analysis of the effects of a police ride-along program on the attitudes of minority group youths. Dissertation Abstracts International, 1973 (June), 33(12-B), 6091.

Pugh, G. M. Concepts of the ideal policeman. Unpublished study, University of Alberta, Edmonton, Alberta. 1976.

Rappaport, D. _Diagnostic psychological testing_, _Volumn 1_. New York: Waverly Press, 1945.

Raubenheimer, I. W. Influence of group and situational differences on the applicability of a personnel test. _Journal of Applied Psychology_, 1970, _54_(3), 214-216.

Reiner, H., Setzer, C., & McCormack, J. Another view on psychiatric screening of police recruits. _Public Personnel Review_, 1960, _21_, 41-55.

Reiss, Albert J. Professionalization of the police. In: Brandstather, A. F., & Radelet, L. (Eds.) _Police and community relations_. Beverley Hills: Glencoe, 1968.

Rhead, C., Abrams, A., Trosman, H., & Margolis, P. The psychological assessment of police candidates. _American Journal of Psychiatry_, 1968, _24_(11), 1575-1580.

Roe, A. _Psychology of occupations_. Yew York: John Wiley and Sons, 1956.

Rotter, J. R., & Rafferty, J. E. Incomplete Sentence Blank. New York: The Psychological Corporation, 1950.

Schmitt, N. Interrater agreement in dimensionality and combination of assesment center judgments. Journal of Applied Psychology, 1977, 2(2), 171-176.

Searles, J. A. & Leonard, J. M. Experiments in the mental testing of Detroit policemen. Detroit Bureau of Government Research, 1936, 141, 54.

Shearing, C. D. & Leon, J. S. Reconsidering the police role: A challenge to a popular conception. Canadian Journal of Criminology and Corrections, 1977, October, 331-345.

Silver, Allan The demand for order in civil society. In: Bordu, D. J. (Ed.) The police. New York: Wiley, 1967.

Simon, W. E., Wilde, V., & Cristal, R. M. Psychological needs of professional police personnel. Psychological Reports, 1973, 33, 313-314.

Skolnick, J. R. The police and the urban ghetto. Research contributions of the American Bar Foundation, 1968, 3, 1-29.

Skolnick, J. R. Justice without trial. New York: Wiley, 1966.

Smith, Bruce Police systems in the United States. New York: Harper and Row, 1960.

Smith, D. & Schau, E. Forced choice performance rating for policemen. Proceedings of the 81st Annual Convention of the American Psychological Association, Montreal, Canada, 1973, 8, 993-994.

Smith, W. The law enforcement profession: An incident of high suicide. Omega, 1970, 1, 295-298.

Spaulding, V. V. A study of nurse and police applicants. Delaware State Medical Journal, 1948, 20, 177-178.

Spencer, G. & Nichols, R. A study of Chicago police recruits: Validation of selection procedures, Police Chief, 1971, 38(6), 50-55.

Stirling, J. W. Changes in the role concepts of police officers. International Association of Chiefs of Police, 1972.

Sterne, D. M. Use of the Kuder Preference Record-Personnel with police officers. Journal of Applied Psychology, 1960, 44, 323-324.

Strong, E. K. Manual for Vocational Interest Blank, Stanford: Stanford University Press, 1933.

Strong, E. K., Jr., & Campbell, D. P. Manual for Strong Vocational Interest Blanks. Stanford, California: Stanford University Press, 1966.

Super, D. E., & Crites, J. O. Appraising vocational fitness by means of psychological tests. New York: Harper and Row, 1962.

Teasley, C. E., & Wright, L. The effects of training on police recruit attitudes. Journal of Police Science and Administration, 1973, 1(2), 241-248.

Terman, L. M. The measurement of intelligence. Boston: Houghton Miffin, 1916.

Terman, L. M., & Otis, A. A trial of mental and pedagogical tests in a civil service examination for policemen and firemen. Journal of Applied Psychology, 1917 1, 17-29.

Thorndike, R. L., & Hagan, E. Ten thousand careers. New York: John Wiley, 1959.

Thorndike, R. L. & Hagan, E. Measurement and evaluation in psychology and education. Toronto: Wiley, 1969.

Thurstone, L. L. The intelligence of policemen. Personnel Journal, 1923, 1,64-74.

Thweatt, W. H. Improving police selection on a shoestring budget. The Police Chief, 1972, 60-68.

Toch, H. H., & Schulte, R. Readiness to perceive violence as a result of police training. British Journal of Psychology, 1961, 52(4), 389-393.

Trojanowicz, R. C. & Dixon, S. Criminal justice and the community. Englewood Cliffs, New Jersey: Prentice-Hall, 1974.

Turner, W. W. The police establishment. New York: Putnam, 1968.

Tversky, A. Intransitivity of preferences. Psychological Review, 1969, 76, 31-48.

Vollmer, A. Police bureau survey. City of Portland Oregon, 1947.

War Department, Army Group Examination Alpha. 1919.

War Department, Adjutant General's Office, Personnel Research Section. The Army General Classification Test: with special reference to the construction and standardization of forms 1a and 1b. Journal of Educational Psychology, 1947, 38, 385-420.

Webster encyclopedic dictionary of the English language. Thatcher, V. S. & McQueen, A. M. (Eds.), Chicago: Consolidated Books Publishers, 1973.

Wechsler, D. Wechsler-Bellevue Intelligence Scale. New York: Psychological Corporation, 1946.

Wechsler, D. Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Weider, A., Wolff, H. G., Brodman, K., Mittelman, B., & Wechsler, D. Cornell Index. New York: Psychological Corporation, 1948.

Weiner, N. L. The role of the police in urban society: Conflict and consequences. Indianapolis: Bobbs-Merril, 1976.

Weislogel, R. L. Development of situational tests for military personnel. Personnel Psychology, 1954, 7, 492-497.

Westley, W. The police: A sociological study of law, custom and morality. PhD dissertation, Department of Sociology, University of Chicago, 1951.

Westley, W. A. Violence and the police. Boston: Colonial Press, 1970.

Wilson, J. Q. A national survey of police and community relations. East Lansing: Michigan State University, 1967.

Wilson, J. Q. Varieties of police behavior. New York: Atheneum, 1970.

Wilson, J. Q. Varieties of police behavior. Cambridge: Harvard University Press, 1968.

Wolfgang, Marvin E. The police and their problems. In: Brandstather, A. F. & Radelet, L. A. (eds.) Police and community relations: A source book. Beverley Hills: Glencoe, 1968.

Wonderlic, E. F. Wonderlic Personnel Test. Northfield: Illionis, 1942.

Wonderlic, E. F. Wonderlic Personnel Test manual. Northfield: Illinois, 1961.

Wonderlic, E. F. A cooperative research study of minimum occupational scores for the Wonderlic Personnel Test. Northfield: Illinois, E. F. Wonderlic & Associates, 1966.

Yinger, J.M. Toward a field theory of behavior. New York: McGraw-Hill, 1965.

Zaice, J. E. Measured interests, personality, and intelligence of professional policemen. Unpublished master's thesis, Washington State University, 1962.

Zurcher, J. C., Miller, D., Rounds, J. C. Selecting effective police sergeants. The Police Chief, 1975, January, 28-30.

APPENDIX 1

Job Descriptions

Summaries of Job Descriptions for Policemen

(1) Accurate memory and observation.

(2) Reasonability.

(3) Analytical judgment.

(4) Ability to follow directions.

(5) Ability to organize material.

(6) Mental alertness.

(7) Speed of decision.

(8) Judgment (commonsense).

(9) Determination.

(10) Social intelligence.

(11) Aggressiveness.

(Holmes, 1942, p. 557)

This above job description emphasized mental function rather than interpersonal qualities; its focus is on organization and efficiency in dealing with criminal acts. The activities of a policeman from this point of view appears to be that of an efficient detector, preventer, or solver of crimes.

In a later job description developed for the Chicago police department (Baehr, Furcon, and Froemel, 1968) more interpersonal skills and maturity were emphasized. The twenty qualities of this job description were:

(1) Endure long periods of monotony and yet react quickly and effectively to problem situations.

(2) Gain knowledge of people in patrol area.

(3) Exhibit initiative, problem solving capacity, effective judgment and imagination in coping with complex situations.

(4) Make prompt and effective decisions.

(5) Demonstrate mature judgment in deciding to make an arrest, give a warning or use force.

(6) Be critically aware of signs of unusual conditions or circumstances.

(7) Exhibit a number of complex psychomotor skills.

(8) Adequately perform communication and record keeping skills.

(9) Have facility to act effectively in extremely divergent interpersonal situations.

(10) Endure verbal and physical abuse from citizens and offenders.

(11) Exhibit a professional, self-assured presence and a self-confident manner.

(12) Be capable of restoring equilibrium to social groups.

(13) Be skillful in questioning offenders, victims and witnesses.

(14) Take charge of situations yet not unduly alienate participants or bystanders.

(15) Be flexible enough to work under loose supervision in some situations and under direct supervision in other situations.

(16) Tolerate stress in a multitude of forms.

(17) Exhibit personal courage in the face of dangerous situations.

(18) Maintain objectivity while dealing with a wide variety of people and groups.

> (19) Maintain a balanced perspective in the face of constant exposure to the worst side of human nature.
>
> (20) Exhibit a high level of personal integrity and ethical conduct.
>
> (Furcon, 1972, pp. 5-6.)

The qualities contained in this above job description suggest good personal development and maturity.

APPENDIX 2

Attitudes of Edmonton Residents towards Policemen

Appendix 2

Attitudes of Edmonton Residents
Towards Policemen (1976)

	Unfavorable%	Favorable%
Feelings toward police	06.0	94.0

	"Bad"	"Good"
Overall Job Performance	2.0	98.0
Police Public Relations Rating	5.9	94.1
Police behavior		

	Yes%	No%
Ever treated impolitely	24.9	75.1
Ever seen impolite treatment	31.4	68.6
Ever treated unfairly	17.1	82.9
Ever seen unfair treatment	22.9	77.1
Ever physically mistreated	03.4	96.6
Ever seen physical mistreatment	15.7	84.3
Ever seen police neglect	20.6	79.4

	Never	Hardly Ever	Sometimes	Most of the time	Always
Honesty	29.9	49.5	18.4	01.6	00.6
Courteous	00.8	01.3	12.4	63.5	21.9
Helpful	35.4	39.8	19.1	03.5	02.2
Fair	00.5	01.8	09.2	74.7	13.7
Competent	01.1	02.0	11.9	73.2	11.8
Efficiency	09.3	39.9	44.6	05.3	01.0

	--Dissatisfied--			--Satisfied--		
	Very	Moderate	Slight	Slight	Moderate	Very
Treatment of General Public	00.7	01.3	04.0	05.5	45.7	42.8
Prompt in response to calls	02.5	02.7	06.3	06.6	38.4	43.5
Enforcing the Law	01.0	01.7	06.5	10.1	53.3	27.5
Traffic Control	02.4	04.7	10.3	12.7	42.1	27.8
Treatment of Law Breakers	05.6	05.0	10.4	11.0	44.0	24.0

APPENDIX 3

Police Studies: Criteria Used

Appendix 3

Police Studies: Criteria Used

STUDY	CRITERIA
Martin (1923)	Combined rating and ranking by four supervisors
Thurstone (1923)	High service/low service record
Ostrander (1925)	Supervisors' ratings
Merrill (1926-27)	Chief's ratings
Berman, et al. (1934)	Efficiency ratings by captain (above average, average, below average)
Diehl et al. (1933)	Ratings by supervisors
Searles & Leonard (1936)	Dismissal vs. stay
Humm & Humm (1950)	Dismissal vs. stay
Cross & Hammond (1951)	Dismissal vs. stay one year
Dubois & Watson (1954)	Training school marks, achievement test, marksmenship, 10 week service rating
Mullineaux (1955)	Marks in academy
Marsh (1962)	Job performance ratings from many who had supervised and observed
Blum (1964)	Background investigation, serious police misconduct
Colarelli & Siegel (1964)	Eight job performance variables: moving hazardous arrest, moving hazardous warning, other arrests, services rendered, light corrections, miles per contact without radar, hours perarrest

Appendix 3 (continued)

Police Studies: Criteria Used

Morman, et al. (1965)	Judged by 5 supervisors on 5 TAV variables
McConnel (1967)	Rating by supervisors
Hankey (1968)	Academy marks, performance rating, commendations
Baehr et al. (1968)	Paired comparison, supervisors' ratings, departmental performance, tenure, awards, complaints, disciplinary action, arrests, times absent
Mills, McDevitt & Tonkin (1966)	Training class marks
Levy (1967)	Success, failure, terminators
McAllister (1970)	Lost time, time lost for injuries, completion of training, recognition of outstanding performance, absence of disciplinary charges, completion of probationary period, evaluation by supervisors
Leiren, Kiker, & Phelan (1971)	Absences, commendations, accidents, supervisors' ratings (emotional stability, appearance, total impression), training class marks
Hooke & Krauss (1971)	Successful, unsuccessful
Spencer & Nichols (1971)	FTQ (failure to qualify based on background investigation), patrolman effectiveness (police department ratings, number of awards, number of index crime arrests, sustained complaints, lost time from work, sick leave abusers' list, safety record)
Hogan (1971)	Staff ratings, performance ratings
Thweatt (1972)	Drop out vs. stay

Appendix 3 (continued)

Police Studies: Criteria Used

Smith & Schau (1973)	Peer nomination, paired comparison, graphic ratings
Zurcher, Miller & Rounds (1975)	In-basket test, paired comparison by supervisors, by sergeants
Hanley (1975)	Forced choice evaluation

APPENDIX 4

Incomplete Sentence Blank

PREVIOUSLY COPYRIGHTED MATERIAL,

IN APPENDIX 4, LEAVES 256-258,

NOT MICROFILMED.

INCOMPLETE SENTENCES BLANK - ADULT FORM, by Julian B. Rotter.
Copyrighted 1950 by The Psychological Corporation, 304 East
45th Street, New York, N.Y., U.S.A. 10017.

APPENDIX 5

Scoring Form for the Incomplete Sentence Blank

SENTENCE-COMPLETION RESPONSE RATING

Subject ________________

Directions: Study the categories where the response might be placed. Read the completion items one at a time. Make a decision and enter the number of the item in the space provided. Be sure to place every response in a category. Many responses will be borderline. When in doubt "force" a choice. After all the items have been classified rate (✓) the Level of Adjustment (Scale 1 to 5, lowest to highest).

		Totals
Harmless Generalization		___
Harmful Generalization		___
Contribution to Society		___
Destructive to Society		___
Feelings of Adequacy		___
Feelings of Inadequacy		___
Liking for Others		___
Disliking for Others		___
Liking: Other than Human		___
Disliking: Other than Human		___
Others Like Subject		___
Others Dislike Subject		___
Self-Serving		___
Response Omitted		___

OVERALL RATING (Rater____, Date____)

Level of Adjustment

Liabilities Far Outweigh Assets	Many Problems	A Few Problems	Fairly Good Adjustment	Excellent Adjustment

Code M = Mature; Mp = Maturity Potential; CrM = Creative Mature; Hm = humor; Fx = Fixation; N = Neurotic;
Sp = Sociopathic; P = Psychotic; D = Defensive; A = Anxiety; G = Guilt; Im = Immature

APPENDIX 6

Scoring Form for the Draw-a-Person

DAP RATING METHOD

Subject:

Date:

Rater:

Rater: Following are a number of scales for rating drawings.
Consider the "sense" of each point on a scale according to the descriptions.
Examine the drawing, referring to one scale at a time.
Encircle the sample characteristics that relate to the drawing. If necessary, write in a characteristic and encircle it.
Make a judgment and mark the scale point score with a ✓.
Enter the score in the space at the right of the scale.
Go to the next scale and rate as before. If a region cannot be rated indicate NR (not rateable).

Total the scores separately under the headings provided and rate the drawing for Overall Level of Adjustment.

SIZE ___

1	2	3	4	5
Huge, over 9" Runs off page Tiny, at 1"	Approximates 3" Mostly between 7-9"		Approximates 3-7"	All in 3-7"

PLACEMENT ___

1	2	3	4	5
Lower right Lower left	Upper right Upper left		Centre Vertical (approximate)	Centre Vertical (precise)

SHEET POSITION ___

1	2	3	4	5
Angular axis Various axes Folded, torn	Horizontal axis		Vertical axis	

AGE OF DP vs SUBJECT ___

1	2	3	4	5
Clearly childhood or old age	Clearly older or younger 5-25 yrs		Approximately same	Slightly older (to 5 yrs)

SEX OF DP vs SUBJECT ___

1	2	3	4	5
Opposite, and parent, child figure Sex not clear Feminine treatment of male	Opposite, Near same age		Same sex	Same, and self-portrait (rare)

ERRORS ___

1	2	3	4	5
Omissions Many errors No corrections Much reworking False starts	Many errors Many erasures Corrections mediocre		Few evident errors (be generous) Few erasures Good corrections	No errors, or none evident (rare)

ARTISTIC QUALITY ___

1	2	3	4	5
Gross disproportion Scribbling Poor planning Transparancy Diagrammatic Stick drawing	Male, female drawn alike Kindergarten effort		Fair drawing (be generous)	Very well planned and drawn

IDENTITY (Clothing) ___

1	2	3	4	5
Antisocial Hippie, hobo. Another century, planet Cartoon, robot Skeleton, nude	Dude, flashy Sports Priest, nun Cowboy, clown Sex object Jewels, buttons, pockets		Presentable most places Conventional Job or leisure suggested	Presentable anywhere Smart-looking Profession or social role suggested Conservative

1	2	3	4	5 LINE QUALITY
Figure unclear Tremulous Smudged Pressure tears paper	Light, faint Broken lines Heavy pressure Sizeable erasures		Firm, but sketchy in places	Firm, confident Few breaks Consistent pressure

1	2	3	4	5 PROPORTION
Very exaggerated in any region	Somewhat unreal		Fair proportions (be generous)	Excellent proportions Like a real person

1	2	3	4	5 MENTAL DEVELOPMENT
Sub average Circles, squares Scribbles, stick No details	Low average Crude Many details absent		Average Most details present	Clearly above average All details in and appropriate Imagination shown

1	2	3	4	5 STANCE
Toppling Feeble, reclining Defiant Legs crossed	"Posed" Feet wide apart, tight togethor Sitting, bending		Good, natural, but slightly stiff	Good, natural, easy

1	2	3	4	5 FACING POSITION
Back view Upside down Face between legs	Side view body entire DP		Front view body only	Front view entire DP

1	2	3	4	5 ACTION
Aggressive Submissive Death scene Running, except sports Coitus, masturbation Elimination Drugs, alcohol Uncoordinated	Just standing, sitting Smoking, eating Private; prayer, shaving		Sublimation, self-serving	Sublimation, others-serving, constructive

1	2	3	4	5 PARTNER
Unsuitable by age or sex Sensual treatment	Unsuitable by age Nearly nude, suggestive Glamourized		Suitable partner Not overdone	Suitable partner attractive sublimated

1	2	3	4	5 MOUTH
Omitted Exaggerated Teeth shown, if not smiling Laughing Angry, menacing Sneering Mustache overdone	Thick lips Feminized on male Sensual on female Slash line Depressed, down Cigarette, pipe Erasures		Conventional Including mustache(neat) on male	Smiling, pleasant Mustache (neat) on male

1	2	3	4	5 NECK AND CHIN
Omitted Very thick Covered by long beard Stick	Thin, long Strain lines Double chin "Choke" collar or necklace Tie in motion		Slender, normal Open collar Beard (neat) on male	Strong, normal (male or female) Ascot, scarf Clean shaven

1	2	3	4	5 CHEST
Extreme size Back view Obese, female breasts on man Internal organs shown Stick	Exaggerated Narrow, thin Fat, muscles shown Breast emphasis on female Profile Erasures		Good proportion Dressed for leisure or athletics	Good proportion Presentable street dress

1	2	3	4	5 WAIST
Beer belly, extreme Very tiny, "wasp" Internal organs shown	Protruding abdomen Flabby, "spare tire" Muscles shown Suggestive dress Enlarged belt, buckle Erasures		Fair proportion	Slim, good proportion Presentable street dress

1	2	3	4	5 GENITAL REGION
Omission of region is conspicuous, (rest of body, hips, legs drawn) Size exaggerated Nude genitals	Clothing, but genitals suggested Drawing stopped at waist Buttons at fly emphasized Shading, profile, erasures		Clothing, but sketchy lines, delicate treatment	Clothing, firm lines

1	2	3	4	5 HIPS
Feminized, curved on male Nude buttocks Stick	Enlarged Narrow Profile, flat Shading Erasures		Good proportion Sketchy lines Presentable street dress	Good proportion Firm lines Presentable street dress

1	2	3	4	5 LEGS
Omitted Extreme size Stick Deformity, crutches, cane Absurd attachment Incorrect number	Exaggerated Thin, frail Lopsided, short Muscles shown Poor attachment Joints emphasized Shorts, briefs worn Shading, erasures		Adequate, good proportion Presentable street dress Realistic attachment	Solid, strong, but not overdone Good proportion Presentable street dress Realistic attachment

1	2	3	4	5 FEET
Omitted Exaggerated Stick Spiked, pointed Deformed Jackboots, decorated Kicking action, except sports	Rather large Too small On toes Barefoot, stockings, sandals Turned to side (90°), or in Toes emphasized Unfinished, incomplete Shaded, erasures		Well-drawn, good proportion Good shoes	Well-drawn, good proportion Solid, firm Good shoes Strong ankles Walking motion

1	2	3	4	5 GROUND LINE
Water, mud, snow, rocks, sand, tall grass Scaffolding Clouds, tightrope	No line		Just a line	Actual ground portrayed. Street floor, lawn

NOSE

1	2 3	4	5
Omitted Exaggerated Broken, bandaged Nostrils enlarged "Alcoholic"	Very large Very simplified Small, cute, sharp Profile, erasures	Normal	Well-drawn, normal

EYES

1	2 3	4	5
Omitted Only one in full face Heavily made up Slits, points, crosses Long lashes on male Crossed, blind, mask, patch, sick	Large, small Shades worn Circles, no pupils Erasures	Well-drawn, normal including spectacles	Well-drawn, normal Proportionate Commanding Friendly

HEAD (Hair,Ears)

1	2 3	4	5
Omitted Very large head Dishevelled, wayout frizzy hair. Very long on male Bald Ears very large Earrings on male	Head too small or outsize, tilted Messy, undone hair "Model" perfect Ears fairly large, uneven Profile Erasures	Good proportion Neat, fashionable hair Ears appropriate	Good proportion Neat, fashionable Shaded, attractive Ears appropriate

FACIAL EXPRESSION

1	2 3	4	5
Omitted features Menacing, silly Dehumanized, sick Double profile Shading (unless art)	Exaggerated features Unhappy, wrinkled Exhausted Profile, erasures	Well-drawn serious	Well-drawn pleasant

SHOULDERS

1	2 3	4	5
Extreme size Huge muscles Stick	Somewhat exaggerated Muscles shown Narrow, thin, square Lopsided, joints emphasized Erasures	Good proportion Dressed for leisure, athletics	Good proportion Presentable street dress

HANDS

1	2 3	4	5
Omitted, hidden Exaggerated Talon, spike fingers Fists, weapon held Deformed, bandaged Touching face, or genital region Incorrect number of fingers Stick	Small, mitten, petal, baby Unusual nails Gloves, unless winter garb Many rings Shaded (unless art) Badly drawn, uneven, unfinished Erasures	Well-drawn Large	Well-drawn Large Constructive action

ARMS

1	3	4	5
Omitted Extreme size Very short, stick Tattoos, bandaged Deformity, atrophy Incorrect number, absurd attachment	Exaggerated Muscles shown Thin, weak, lopsided Joints emphasized Poor attachment Erasures	Proportionate to shoulder Evenly sized Realistic attachment	Proportionate to shoulder and body Realistic attachment Constructive action

Scoring Summary

Overall Total ________

Overall Weighted Total ________

Rate the drawing for Overall Level of Adjustment.

	1	2	3	4	5
Guiding Scale (Approximate)	0.0	38.75	77.50	116.25	155.00

APPENDIX 7

Scoring Form for the Thematic Apperception Test

TAT Story Rating Method Subject ____________ Rater ______ Date ______

Directions: Study the categories where the story content might be placed. Reduce the story to its simplest terms. Enter the significant aspects in the spaces provided. Extract the Credo for each story and rate its worth for real life behavior (1 is lowest, 5 is highest). Add the ratings for a total score.

Category	Card	Card	Card	Card	Card	Card	Card	Action
Problem set by Other(s)								
Problem set by Self								
Feelings of Adequacy								
Feelings of Inadequacy								
Solution by Other(s)								
Solution by Self								
Good Solution to Problem								
Poor Solution to Problem								
Liking for Other(s)								
Disliking for Other(s)								
Story Incomplete, None								
Credo								Rating Total
Rating: Circle a number	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	

APPENDIX 8

Scoring Form for the Situation Test and the Three-Minute-Test

3MT RATING Subject_______ Date_______ Rater_______

Directions: Study the categories and scales used for evaluation of performance. Note the descriptions of high (5) and low (1) points. Review the applicant's videotape. Mark the presentation (3MT) that is being considered. Rate the applicant's performance on the general scales. Where a category does not apply, write N/A. Finally, on the basis of the general evaluation, rate Level of Development.

People_____ World_____

Introduction	Worried What to say?	1 . . . . 2 . . . . 3 . . . . 4 5	Clever start Catches interest
Ideas	Dull Nothing New	1 . . . . 2 . . . . 3 . . . . 4 5	Fresh Great new Ideas
Clarity	What was said?	1 . . . . 2 . . . . 3 . . . . 4 5	Clear as crystal
Voice	Very quiet, low Very loud	1 . . . . 2 . . . . 3 . . . . 4 5	Well modulated Good control
Organization	All over the place	1 . . . . 2 . . . . 3 . . . . 4 5	Good sequence
Reasoning	Nonsense	1 . . . . 2 . . . . 3 . . . . 4 5	Very logical
Closure	Stops mid-flight Runs out of ideas	1 . . . . 2 . . . . 3 . . . . 4 5	Closes with punch Good timing
Poise	Hesitant, nervous Feels threatened	1 . . . . 2 . . . . 3 . . . . 4 5	Confidently says what wants to say
Quality of Content	Attitude very negative Destructive thoughts Pessimistic, cynical	1 . . . . 2 . . . . 3 . . . . 4 5	Attitude very positive Constructive thoughts Optimistic, has faith
Listener's Reaction	Bored, turned off Disgusted, shakes head	1 . . . . 2 . . . . 3 . . . . 4 5	Delighted, applause Encore!

Overall Rating

Level of Development	Immature thinking Lacks communication skills	1 . . . . 2 . . . . 3 . . . . 4 5	Mature thinking Good communication skills

SITEST RATING METHOD Subject______ Date______ Rater______

Directions: Study the categories and scales used for evaluation of performance. Note the descriptions of high (5) and low (1) points. Review the applicant's videotape. Mark the situation that is being considered (✓). Rate the applicant's performance immediately on the six general scales. Where a category does not apply, write N/A. Finally, on the basis of the general evaluation, rate Level of Development.

Launchair____ Directions____ Lifeguard____ Dog____

Approach to Situation 1 . . . 2 . . . 3 . . . 4 . . . 5
Too timid / Too authoritative — Smooth, Confident / Friendly

Clarity of Verbal Request or Response 1 . . . 2 . . . 3 . . . 4 . . . 5
Unclear communication / Confused, Mumbled — Clear Message / Brief, Well-spoken

Consideration Given to Other's Concerns 1 . . . 2 . . . 3 . . . 4 . . . 5
Does not listen / Impatient, Indifferent — Listens with interest / Patient, sympathetic

Effectiveness of Solution to Problem 1 . . . 2 . . . 3 . . . 4 . . . 5
No real solution / Solution resisted — Good solution / Accepted as Helpful

Confidence Shown in Handling Situation 1 . . . 2 . . . 3 . . . 4 . . . 5
Does not know what to do / Insecure, stiff, nervous — Knows what to do / Takes charge, unstressed

Interpersonal Skills 1 . . . 2 . . . 3 . . . 4 . . . 5
Poor Handling / Makes an Enemy — Wise Handling / Makes a Friend

---

Overall Rating

Level of Development 1 . . . 2 . . . 3 . . . 4 . . . 5
Immature / Lacks Resources — Mature / Has Good Resources

APPENDIX 9

Instructions and Rating Form for Performance Assessment

TO: SERGEANTS WITH MEMBERS OF THE RESEARCH GROUP

FROM: PERSONNEL SECTION

Within your squad are police constables who are part of a research project concerned with developing a battery of selection tests that will predict later performance on the job. Due to the nature of the project two additional measures of job performance are required.

(1) a ranking measure

(2) a rating of OVERALL PERFORMANCE

The ranking measure will give the constables position in the squad compared with the others; the OVERALL PERFORMANCE rating will give a numerical figure to the judgment of actual performance. You will be supplied with forms for each squad member.

When you receive and examine these forms:

(1) Arrange the forms in order of merit on OVERALL PERFORMANCE. This will give you a ranking of constables in your squad from highest to lowest. Enter a number to indicate ranking position, assigning number "1" to the highest ranked, number "2" to the second highest ranked, and so on.

(2) The forms are now ready for rating on OVERALL PERFORMANCE. A suggested method is to lay out the forms in such a manner that they are placed in order from highest to lowest. This placement will afford you opportunity for visual review before the rating of OVERALL PERFORMANCE. Taking note of the captions under the rating scale decide how the top member is to be rated and enter your rating. Follow the same procedures for the lowest ranked member. After this, rate the others. Please confine your rating to the scale demarkations.

Reg. No. ____ Constable ______________________

Ranking in squad ________

Total Number of men in squad ________

OVERALL PERFORMANCE RATING (✓)

1	2	3	4	5	6	7	8	9
Very Poor	Poor	Fair	Below Average	Average	Above Average	Good	Very Good	Outstanding

APPENDIX 10

Problems with Reliability of Performance Ratings

Problems with Reliability of Performance Ratings

Reliability problems often occur for studies that attempt to measure the job performance of policemen. The focus of this problem involves obtaining more than one performance rating for each subject. For this research, the subject's job performance was assessed on two job performance measures (ranking and rating) by his squad sergeant. The reliability measure for the initial performance rating was based on the correlation between these two measures. This type of reliability indicator was not considered adequate because it only measured reliability within one rater and not between raters. (This would allow for one rater to consistently give poor ratings without being detected.)

In considering this problem further a review of solutions to criterion problems in other research was undertaken. Raubenheimer (1970) presented a study that attempted to investigate the feasibility of applying an American developed personnel test (Purdue Clerical Adaptability Test) to workers in the Republic of South Africa. The reliability of the criterion was presented as follows.

> The estimation of criterion reliability, however, posed a different problem. A test-retest procedure was rather impossible since the paired comparison ratings could be obtained only once. Furthermore only one judge per work group (the supervisor) was available to rate the employees and an interrrater reliability estimate could not be obtained. On the other hand, a wealth of evidence endorses the effectiveness of the paired comparison system as a rating technique. Studies, for example, by Bartlett, Hermann, and Rettig (1960) and Lawshe, Kephart, and McCormick (1949) indicate that this technique can be regarded as superior to all other techniques in several ways---including interrater reliability (p. 215)

The problems faced by Raubenheimer (1970) were similar to that of the present research. Only one supervisor per group existed in both research projects and also the rating approachs were comparable (paired comparison and rank ordering). His research, however, is obviously weakened by the poor reliability measures. The main deficiency within Rauberhiemer's research is that he suggested that a method proven reliable in one culture would also be reliable in another. This assumption of cross cultural applicability, may be difficult to accept.

The principle underlying this solution, however, is an interesting one. It suggests that the reliability of a measuring instrument can be proven independently of the research group. This principle is well established for a psychological test's reliability. A test is proven reliable and then given to different groups with this reliability measure accepted. This principal has been applied by Bruch (1977) who attempted to predict the need for counseling with the use of the Psychological Screening Inventory (PSI). Ratings of "social adjustment" were needed for this research. "A total of 41 upper class residence hall advisers rated all freshmen and sophomores under their supervision on a 10 item behavioral rating scale ...." (p. 239). The reliability problem was dealt with independently of the data collection.

> "Prior to collecting ratings, estimates of interrater reliability were obtained. Three female raters and two male raters, selected at random from the group of 47, assessed, respectively, a group of women students and a group of male students, all of whom were separate from the current student sample" (p. 239).

The correlations among raters ranged from r=+.72 to r=+.89. For the actual research project the following occurred.

> No interrater reliability estimates were taken for the ratings of the current sample, since each of the 41 raters evaluated different individuals. (p. 239)

For Bruch's research, "social adjustment" ratings were shown to be reliable independently of the data collection. For this present research project on policemen the overall job performance measures could also be proven reliable independently of the data collection. The reliability results could be retrospectively applied to the present research.

The best approach to proving the reliability independently would involve having all raters assess all ratees. This approach has been demonstrated by Schmitt (1977). In his study he attempted to examine the interrater reliability of assessment center ratings. Four raters assessed 101 potential middle managers on 17 dimensionalities. Interrater reliabilities (before a discussion of ratings) ranged from r=+.52 to r=+.88. These ideal conditions are often not possible in the actual work environment. For the present research few sergeants were familiar with more then one squad. These conditions preclude the possibilites of using the ideal approach to gathering reliability information.

Solutions to problems similar to this have occurred in previous research. For example Campbell, Dunnette, Arrey, and Hellervik (1973) developed a rating scale for department managers within large stores. Their relaibility problem was handled as follows.

> Each department manager was rated by both his store manager and assistant store manager using both the scaled expectation method and the summed rating procedures. (p. 17)

In this situation all raters could not rate all ratees. The correlation between pairs of raters on ratees was considered as an acceptable reliability indicator.

A similar approach where all the raters did not rate all the ratees was used by Motowidlo and Borman (1977). Their research attempted to measure the morale of 47 platoons in the U.S. Army (in foreign lands). The interrater reliability measure of platoon morale was handled as follows.

> With the newly developed scales, morale of each platoon was rated independently by two persons, the company commander and the first sergeant in the company of which the platoon was a part. (p. 179)

Interrater reliabilities ranged from r=+.47 to r=+.75. In this situation the company commander may have rated more than one platoon while the sergeants involved only rated their own platoon. The reliability was based on pairs of raters, one of the raters in the pair possibly being involved in more than one pair of raters.

Landy, Farr, Saal, and Freytag (1976) had a similar approach to obtaining interrater reliability. Their research was directed towards the development of a rating scale for police officers. For the purposes of selecting different performance levels of policemen, constables were rated on "overall worth" to the police department. "Where possible pairs of raters were identified who could provide parallel ratings for a single ratee. These pairs were used as a basis of scale-reliability analyses" (p. 275). Median reliability correlations ranged from r=+.47 to r=+.74. In this research of Landy and colleagues (1976), it appears that when two raters happened to know the same ratee the shared knowledge was used as an indicator of interrater reliability. This approach, although not comprehensive, is probably still representative of interrater agreement.

A comparable approach in the present research could only be used if one sergeant knew another sergeant's entire squad. This occurrence (for the present research) would be necessary to indicate the reliability of the rank-ordering measure. Within this police department, sergeants only infrequently change squads. Sergeants are not in a good position to rate the job performance of constables not in their squad. An approach similar to Landy and others (1976) therefore appears difficult to implement for this research. A modification of the method of using rater pair correlations plus completing the reliability measures independently of the research may, however, be possible.

It is suggested for this research that the following approach be used to better indicate the reliability of the job performance measures. Rating pairs should be found to rate each squad. Although relatively few sergeants know all of the constables in another squad, a senior constable in each squad (who is awaiting promotion to sergeant) could be asked to assess squad members with the rating methods used in this research. Also staff-sergeants may be able to rank some of the squads that they supervise. From these rating pairs (the sergeant and senior constable for each squad, the sergeant and staff-sergeant) interrater agreement measures could be gathered.

These ratings (rank-order in squad, overall job performance) could be obtained by the pairs of raters during the present time period (after the intial data were collected). The correlations between the rating pairs would indicate independently the degree of the reliability of these job performance measures. These reliability figures could be retrospecively applied to the present research.

Proposed Action for Research

(1) Conduct another ranking and rating of the subjects in the research that involves job performance appraisals from squad sergeants, senior constables, and staff-sergeants. Inter-rater reliability scores could then be obtained.

(2) As a supplementary measure of reliability between squads, all raters, after rating and ranking their squad, will be asked to rate and rank the other subjects in this study. Part of this task will also include an indication of "how well known" the subject is. This will allow staff-sergeants, sergeants, and senior constables from one squad to assess the job performnce of men in another squad. The analyses of these data may involve a partial correlation where the "familiarity" variance is subtracted from the total variance so that the between rater correlations more accurately reflect a reliability measure. An example of the form used for this supplementary measure follows.

Name

of Rater _______________ Date __________

Performance Rating of

Men not in your Squad

(1) Rate the constables in terms of how well you know them by placing an X in one of the "Familiarity" columns.

(2) Strike out the names of the constables you "Don't Know" or "Know Slightly".

(3) Rank order the remaining constables in terms of "overall job performance" as if they were in your present squad.

(4) Rate these same constables in terms of "overall job performance" on the 1-9 scale in the same manner as you have rated the men in your squad.

Constable	Familiarity				Rank	Rating
	Don't Know	Know Slightly	Know Well	Know Very Well		Overall Performance
Smith, A. (000)						123456789
Smith, B. (000)						123456789
Smith, C. (000)						123456789
Smith, D. (000)						123456789
Smith, E. (000)						123456789

APPENDIX 11

Ratings by Sergeants of Three Written Incidents

Name ____________________ Date ______________

Please rate the following situations in terms of the constable's "overall performance".

Constable A approaches a drunk on the street, who is extremely intoxicated (staggering and falling down), is dirty and wearing ragged clothing.
The constable says: "Hey you. Your're just another drunken pig, aren't you? I've seen a thousand like you and you're not worth dirt."
The drunk says: "But officer, I'm not that drunk".
The constable says: "For your own safety, I'll have you transported into the custody of the police department until you're sober." At this point the paddy wagon is called and the man is taken away.

OVERALL PERFORMANCE RATING ( )

1 2 3 4 5 6 7 8 9
├─:─┼─:─┼─:─┼─:─┼─:─┼─:─┼─:─┼─:─┤

Constable B has just followed a car doing 50 in a 30 zone. He pulls the car over to the curb and approaches the driver. He says: "What right do you have to travel at that speed? Give me your documents and hurry it up." This was followed by a traffic ticket being issued.

OVERALL PERFORMANCE RATING ( )

1 2 3 4 5 6 7 8 9
├─:─┼─:─┼─:─┼─:─┼─:─┼─:─┼─:─┼─:─┤

Constable C has been called to a family dispute as a result of the wife's complaint. On the doorstep the husband answers and tells the constable to leave.
The constable says: "A Mrs. Smith phoned in a complaint. May I speak with her please."
Mrs. Smith enters the doorway and the constable says: "Mrs. Smith, I am responding to the complaint you phoned to the police department. Could I step in and discuss this matter with you?."
She agrees. The constable says upon entering the house addressing both the man and the woman: "I would like to speak to both of you, but separately, about this complaint." Turning to the man, "I would like to speak to Mrs. Smith first, alone please. I'll hear your side next". The man leaves the room. After hearing both sides of the incident and referring the couple to a community agency the constable departs having defused the situation.

OVERALL PERFORMANCE RATING ( )

1 2 3 4 5 6 7 8 9
|—:—+—:—+—:—+—:—+—:—+—:—+—:—+—:—+—:—|

Table 43

Ratings by Sergeants of the Three Written Incidents

Incidents	Mean	Standard Deviation	Range	Standard Error
Intoxication	2.30	1.55	5	.32
Speeding	2.26	1.48	5	.32
Family Dispute	7.96	.18	3	.18

APPENDIX 12

The Attribute Rating Form given to the sergeants

Name ________________ Date ________________

(1) Rate (by circling) the attributes below in terms of their importance for the "overall performance" of a policeman.

Attributes of Job Behavior	Lesser Importance					Greater Importance				Rank
Tough	1	2	3	4	5	6	7	8	9	
Courteous	1	2	3	4	5	6	7	8	9	
Cold	1	2	3	4	5	6	7	8	9	
Common Sense	1	2	3	4	5	6	7	8	9	
Intelligence	1	2	3	4	5	6	7	8	9	
Friendly	1	2	3	4	5	6	7	8	9	
Take-charge	1	2	3	4	5	6	7	8	9	
Honesty	1	2	3	4	5	6	7	8	9	
Physical	1	2	3	4	5	6	7	8	9	
Kind	1	2	3	4	5	6	7	8	9	
Hard-nosed	1	2	3	4	5	6	7	8	9	
Fairness	1	2	3	4	5	6	7	8	9	
Self-Control	1	2	3	4	5	6	7	8	9	
Alertness	1	2	3	4	5	6	7	8	9	
Ruthless	1	2	3	4	5	6	7	8	9	
Enforcement	1	2	3	4	5	6	7	8	9	
Courage	1	2	3	4	5	6	7	8	9	
Watchman	1	2	3	4	5	6	7	8	9	
Public Servant	1	2	3	4	5	6	7	8	9	
Forceful	1	2	3	4	5	6	7	8	9	

(2) Rank order the attributes (the best being number 1, the second best 2, and so on) in terms of their performance for the "overall performance" of a policeman.